Magued Iskander
Vikram Kapila
Mohammad A. Karim *Editors*

Technological Developments in Education and Automation

Springer

Technological Developments in Education and Automation

Magued Iskander · Vikram Kapila ·
Mohammad A. Karim
Editors

Technological Developments in Education and Automation

Springer

Editors
Magued Iskander
Polytechnic Institute of NYU
Civil Engineering Dept.
333 Jay Street
Brooklyn NY 11201
USA
Iskander@poly.edu

Dr. Vikram Kapila
Polytechnic Institute of NYU
Mechanical Engineering Dept.
333 Jay Street
Brooklyn NY 11201
USA
vkapila@poly.edu

Dr. Mohammad A. Karim
Old Dominion University
4111 Monarch Way
Norfolk VA 23508
USA
mkarim@odu.edu

ISBN 978-90-481-3655-1 e-ISBN 978-90-481-3656-8
DOI 10.1007/978-90-481-3656-8
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009942235

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book includes the proceedings of the 2008 International Conference on Engineering Education, Instructional Technology, Assessment, and E-learning (EIAE 08) and International Conference on Industrial Electronics, Technology & Automation (IETA08). Both conferences were part of the International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CISSE 08). The proceedings are a set of rigorously reviewed world-class manuscripts presenting recent technological developments in education and automation.

CISSE 08 was a high-caliber research conference that was conducted online. CISSE 08 received 948 paper submissions and the final program included 390 accepted papers from 77 countries, representing the six continents. Each paper received at least two reviews, and authors were required to address review comments prior to presentation and publication.

Conducting EIAE 08 online presented a number of unique advantages, as follows:

- All communications between the authors, reviewers, and conference organizing committee were done on line, which permitted a short six week period from the paper submission deadline to the beginning of the conference.
- PowerPoint presentations, final paper manuscripts were available to registrants for three weeks prior to the start of the conference.
- The conference platform allowed live presentations by several presenters from different locations, with the audio and PowerPoint transmitted to attendees throughout the Internet, even on dial up connections. Attendees were able to ask both audio and written questions in a chat room format, and presenters could mark up their slides as they deem fit.
- The live audio presentations were also recorded and distributed to participants along with the power points presentations and paper manuscripts within the conference DVD.

The conference organizers and us are confident that you will find the papers included in this volume interesting and useful. We believe that technology will continue to infuse education thus enriching the educational experience of both students and teachers.

Magued Iskander. PhD, PE
Vikram Kapila, PhD
Mohammad Karim, Ph.D.
June 2009

Table of Contents

Acknowledgements

The 2008 International Conference on Engineering Education, Instructional Technology, Assessment, and E-learning (EIAE) and the resulting proceedings could not have been organized without the assistance of a large number of individuals. EIAE is part of the International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CISSE). CISSE was founded by Professors Tarek Sobh and Khaled Elleithy in 2005, and they set up mechanisms that put it into action. Andrew Rosca wrote the software that allowed conference management, and interaction between the authors and reviewers online. Mr. Tudor Rosca managed the online conference presentation system and was instrumental in ensuring that the event met the highest professional standards. We also want to acknowledge the roles played by Sarosh Patel and Ms. Susan Kristie, our technical and administrative support team.

The technical co-sponsorship provided by the Institute of Electrical and Electronics Engineers (IEEE) and the University of Bridgeport is gratefully appreciated. We would like to express our thanks to Prof. Toshio Fukuda, Chair of the International Advisory Committee and the members of the EIAE and IETA Technical Program Committees.

The excellent contributions of the authors made this world-class document possible. Each paper received two to four reviews. The reviewers worked tirelessly under a tight schedule and their important work is gratefully appreciated. In particular, we want to acknowledge the contributions of the technical committee members who provided the bulk of peer review.

The first and second authors are grateful to the National Science Foundation's GK-12 program grant DGE-0741714 that permitted them to extend their pedagogical interests beyond their own classroom into the k-12 community and share these results through international forums such as EIAE.

Finally, we wish to gratefully acknowledge the support we received from our employers Polytechnic Institute of New York University and Old Dominion University.

Magued Iskander. PhD, PE
Vikram Kapila, PhD
Mohammad Karim, Ph.D.
June 2009

Technical Committee of the 2008 International Conference on Engineering Education, Instructional Technology, Assessment, and E-learning (EIAE 08)

Technical Committee of the 2008 International Conference on Industrial Electronics, Technology & Automation (IETA 08)

Giua Alessandro, University of Cagliari, Italy
Essam Badreddin, University of Mannheim, Germany
John Billingsley, University of Southern Queensland, Australia
Angela Di Febbraro, Politecnico Di Torino, Italy
Aydan Erkmen, Middle East Technical University, Turkey
Navarun Gupta, University of Bridgeport, USA
Junling (Joyce) Hu, University of Bridgeport, USA
Mohamed Kamel, Universityof Waterloo, Canada
Heba A. Hassan, Cairo University, Egypt
Heikki N. Koivo, Helsinki University of Technology, Finland
Larry Hmurcik, University of Bridgeport, USA
Luu Pham, Pitney Bowes, USA
Saeid Nahavandi, School of Engineering & Technology, Australia
ElSayed Orady, University of Michigan, Dearborn, USA
Angel Pobil, Jaume-I University, Spain
Anatoly Sachenko, Ternopil Academy of National Economy, Ukraine
Sadiq M. Sait, King Fahd University, Saudi Arabia
Nariman Sepehri, The University Of Manitoba, Canada
Bruno Siciliano, University of Naples Federico II

Performance-Based Measurement for Thai Educational Organization: A DEA Management Model

Sangchan Kantabutra
School of Management
Mae Fah Luang University, Chiangrai, Thailand
sangchan@mfu.ac.th

Abstract- **This study employs the Data Envelopment Analysis (DEA) technique, an efficiency model, to measure public schools in Thailand. Data came from 455 public upper secondary schools in Northern Thailand. The results reveal that the schools mostly operate inefficiently. To ascertain the efficiency measures, sensitivity analysis reveals the robustness of the resulting efficiency measures on efficiency. Policy to improve school efficiency may thus consider using efficiency-based analysis for measuring school performance to identify which school will yield higher marginal returns to prioritize inefficient schools for treatment, in addition to allocating budget based on block grants or enrolments.**

I. Introduction

With the scope of unstoppable change under globalization age, governments in the Asia-Pacific region are finding it increasingly more difficult to put their educational policies into practices [[1], [2]]. Although spending on education is major government expenditure for the region's countries, the capacity of their educational systems to meet the new demands of this global age is still in question [3].

Thailand, one of the relatively fast-growing economies of Southeast Asia, has been providing the largest share of total public expenditure to education since 1991--even after the 1997 economic crisis began, the government provided 25% of total expenditures, or 4.3% of GDP, for education, which was a greater portion than before, demonstrating its commitment to education for the nation's recovery and development [4]. However, although the Thai government has committed significant resources, its degree of efficiency in providing educational services and procedures is still unclear [4].

This study employs a DEA technique to assess public general upper secondary school efficiency in Northern Thailand. DEA has been recognized as a useful management tool that can be applied for evaluating the relative efficiency of various organizations, e.g., hospitals [5], banks [6], or schools [7], as it compares a given unit to the best-performing units in its peer. The results obtained can identify relatively efficient/inefficient organizations and allow policy makers to develop performance-based policies to assist inefficient ones to improve their performance.

The purpose of this study is to estimate school efficiency and examines the robustness of resulting measures on efficiency by using sensitivity analysis. During the past decade, while most DEA studies have measured school efficiency in the U.S., U.K., Finland, Spain and Australia, very few studies have dealt with schools in Thailand. The novel methodological contribution of this study is a statistical foundation of output efficiencies along with a sensitivity analysis for the robustness on efficiency.

II. The Analytical Framework

The analytical framework of this study would entail: (1) measuring school efficiency, and (2) testing the robustness of resulting on efficiency measures.

The data envelopment analysis (DEA), an efficiency model initially introduced by [8], is used to assess school efficiency, through viewing schools as productive units using multiple inputs to generate multiple outputs. DEA is a deterministic means of constructing a "piece-wise linear" approximation to the frontier based on the available sample. In simple terms, the distribution of sample points is observed and a "kinked" line is constructed around the outside of them, "enveloping'' them (hence the term "data envelopment analysis"). DEA provides a comprehensive analysis of relative efficiency by evaluating each decision-making unit (DMU) and measuring its performance relative to an envelopment surface composed of other DMUs. Each DMU in a given set can then be ranked according to how efficiently it either utilizes its inputs to produce its outputs or it maximizes its outputs using its given inputs, depending on the measuring propose. Units that lie on the surface are deemed "efficient" in DEA terminology, while units that do not lie on the surface are termed "inefficient". The unit of measurement of DMU performance used is an efficiency score. After evaluation of the relative efficiency of the present set of units, the analysis shows how inputs and outputs have to be changed in order to maximize the efficiency of the target DMU. DEA also suggests a benchmark for each inefficient DMU at the level of its individual mix of inputs and outputs.

DEA has been used and cited as an important approach to assess school efficiency internationally [[7], [9], [10], [11],

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_1, © Springer Science+Business Media B.V. 2010

[12]] and has been proposed and assessed Thai school efficiency [[13], [14], [15], [16], [17]] due to its several unique advantages over other traditional techniques, e.g., ratio analysis and regression analysis [18], [19]]. First, DEA can handle multiple inputs and outputs on a simultaneous basis in the input/output education production function. Second, DEA does not require parametric specification of the production function in mathematical form. Third, DEA does not assume behavioral assumptions, e.g., like cost-minimizing or profit-seeking behavior, and this is especially relevant to public schools, which are non-profit organizations. Fourth, inputs and outputs that are contributing to inefficiency are identified and administrators can decide whether a reallocation of resources is necessary or feasible. Finally, managerial strategies for improvement of inefficient decision-making units can be determined.

In the context of public schools, [11] noted that an output maximization DEA model was more appropriate behavioral assumption to take than an input minimization one since school principals should be oriented towards obtaining the best results on the basis of the resources available to them, rather than minimizing these resources, over which they exercise no control –which is a philosophy that is implicit in the output orientation version. A number of studies including [[7], [9], [11], [13], [14], [17], [20], [21]] used the output-oriented approach for assessing school efficiency. The graphical illustration of output maximization DEA model is shown in Figure 1.

Figure 1 illustrates technical efficiency using the case of four schools, A, B, C and D, utilizing a single school input (x) to produce two outputs, y_1 and y_2. The best practice production frontier is defined by the schools that can maximize outputs for a given input. Schools C and D, in this example, are efficient because they lie on the best practice production frontier (isoquant G-G') and the value of the output distance function for each is equal to one. School A and B, which lie inside the frontier, are inefficient and each has an output distance (OA/OA′) and (OB/OB′) of less than one, respectively.

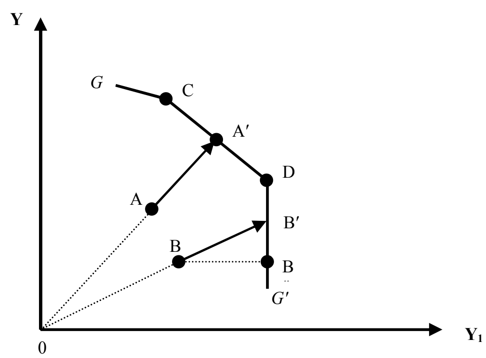


Fig. 1. Output-Oriented DEA Model

DEA technique can be carried out by solving an output-oriented linear programming, incorporating the assumption of constant returns to scale in the following equation [(Eq. (1)].

$$\text{maximize: } Z_0 = \phi_0 + \varepsilon\left[\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right]$$

Subject to:

$$\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ - y_{r0}\phi_0 = 0, \quad r = 1,\ldots,s;$$

$$\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- - x_{i0} = 0, \quad i = 1,\ldots,m;$$

$$s_i^- \geq 0;\ s_r^+ \geq 0;\ \lambda_j \geq 0 \qquad (1)$$

where x_{ij} and y_{rj} represent the observed values of the *i*th input and the *r*th output for school *j* (j = 1, …, n); λ_j is the input and output weights of other schools; s_i^- and s_r^+ are the slacks; ϕ_0 represents the efficiency of the school being evaluated; ε, a small positive constant, guarantees that inputs and outputs are positive and that the slacks do not influence the target function Z_0.

For assessing all schools in the sample, Eq. (1) is *n* times, giving *n* sets of λ_j, one set for each school, to determine its relative efficiency. In an output-oriented DEA model, technical efficiency is measured by the reciprocal of the output distance function [[20], [22]]. The reciprocal of the direct output function gives the proportion by which all outputs can be expanded, given the inputs. This output-oriented DEA model implies that the efficiency score (ϕ) will be equal to or higher than 1. An efficiency score (ϕ) of 1, together with nil values for all the slacks, indicates efficiency, i.e., that the school being evaluated has obtained the maximum possible production on the basis of the limited resources are available to it and the conditions under which it operates. A score of higher than 1 indicates that the school being evaluated could increase its production by the proportion (ϕ-1) without changing its current resources.

Sensitivity analysis was conducted to ascertain the robustness of the efficiency scores from the empirical findings. Two other input and output specifications are conducted to examine whether additional modifications in the educational production function could result in substantive changes in the empirical findings of efficiency measures. In addition to primal specification (Specification 1), one of output variables is decomposed into 3 sub-variables in Specification 2. In Specification 3, one input variable is added. Moreover, it is meaningful to note that there are several factors which may affect the stability of the efficiency results [[9], [10]]. First, the efficiency frontier may be partly based on outlier units or efficient schools that are very different from other units, either genuinely or due to miscoding, measurement error, etc. In such a case, omitting these outliers may change the mean efficiency and rankings based on efficiency scores. Second, the use of different combinations of inputs and outputs may change the ranking of individual schools. Thus, two forms of sensitivity analysis were conducted. First, a form of sensitivity analysis known as "jackknifing" is used to test the robustness of DEA results in regard to outlier schools. Second, Spearman rank correlation between DEA specifications was used to test the changes in their ranking based on efficiency score from one DEA specification to another.

III. DATA AND RESULTS

Data came from 455 public general upper secondary schools that were in Northern Thailand in 2003. All of these schools were under the same regulations of the Office of the Basic Education Commission (OBEC) of the MOE and utilized the same curriculum, thus ensuring homogeneity across the schools, and that all schools had the same production technology.

Inputs (X_i) used here are the teacher-student ratio [[7], [11], [13], [14]] and proportion of students not from low-income families [[9], [13], [14], [23], [24]], to reflect the quantity of resources available (e.g., teachers) and quality of inputs (e.g., students). This study includes SES input in DEA model since it has been argued that if SES was not included in calculating efficiency, the results obtained would not be operationally valid [25]. Because general upper secondary school lasts for three years, input data covers 2001-2003, measured as an average whenever possible. To measure school outputs (y_r), average national test scores [[7], [9], [10], [13], [14]], the number of students who passed their grades after first and second year (average of 2001-02) or were moved up [[10], [13], [14]], and number of graduates [[10], [13], [14]] were included to measure school efficiency. (A full list of descriptive statistics of input and output variables, together with description of variables, is provided in the Appendix.)

Applying the output-oriented DEA model in Equation 1 to each school in the sample yields descriptive statistics of the efficiency measures in Table 1. The average level of efficiency for the sample is 2.2618, indicating that, on average, all outputs being produced could be expanded by 126.18% (2.2618-1) at the given input usage. Variation of efficiency measures is 0.7373 when the gap between the maximum inefficiency measure and best practice is 4.6399 (5.6399-1). This suggests that although efficiency of schools in the set differ from each other by 0.7373 on average, the range between the least efficient and best practice schools in the pooled sample is as high as 4.6399 (5.6399-1). Further, only 4 schools were on the efficiency frontier (with efficiency score of 1). These results indicate that the schools under observation operate efficiently only 0.88% (4/455). In the other words, almost all schools operate inefficiently.

TABLE 1
EFFICIENCY RESULTS

Mean (S.D.)	Maximum Inefficiency Score	Maximum Efficiency Score	No. of Efficient Schools
2.2618 (0.7373)	5.6399	1.0000	4

At the individual school level, the efficiency measure indicates how much output could be proportionally expanded and using the same amount of inputs. For example, in Table 2 the efficiency score of school "Sh17" is 1.6633, implying that all outputs of this school could be expanded by 66.33 % (1.6633 – 1) using the same amount of inputs. In addition, setting adjustment targets for a given school needs to take into consideration that institution's input and output slack. Sh17 has output slack of 42.77 graduates. The output slack figures indicate the extent of further output augmentation after the expansion by 66.33%, while still maintaining the same input level.

On the input side, there is slack of 0.16 SES, representing the potential input reduction for the school even after its outputs have been adjusted for efficiency. Further, adjustment targets for school (Sh17) show that it's passing, graduates, and SAT scores can be raised by 483.55, 191.35, and 23.68 respectively. The SES of this institution can be reduced by 0.16. Slack in the SES index can reflect the school's relative ease of overcoming learning difficulty due to student SES. Similarly, the number of passing and graduates, and SAT of schools are also dependent on the SES.

TABLE 2
ADJUSTMENT TARGETS OF SCHOOL (SH17) IN THE SAMPLE

Variables	Existing Inputs And Outputs	Slacks	Adjustment Targets	Efficient Targets
Inputs				
Teacher-student ratio	3.42	0.00	0.00	3.42
SES	0.80	0.16	0.16	0.64
Outputs				
Passing	729.00	0.00	483.55	1,212.55
Graduates	224.00	42.77	191.35	415.35
SAT scores	35.70	0.00	23.68	59.38

Further, efficiency scores of schools in the sample were calculated using the output-oriented DEA model in Equation 1 together with the two additional specifications. Table 3 compares the efficiency scores of schools the primal specification (1) and the alternative specifications (2 and 3).

TABLE 3
DESCRIPTIVE STATISTICS OF EFFICIENCY SCORES, SPECIFICATIONS 1 – 3

Specification	Mean (S.D.)	Maximum of Inefficiency Score	No. of Efficient Schools	% of Efficient Schools
1*	2.2618 (0.7373)	5.6399	4	0.90
2	2.1017 (0.6649)	5.3242	5	1.10
3	2.0321 (0.6616)	5.0663	12	2.60

* Primal specification

The mean value and number of efficient schools using specification 2, which decomposes SAT scores into verbal, numerical and analytical ability scores in place of total SAT scores in the production model, are higher than those of specification 1. This result may be because schools are likely to be better when they are estimated by various subjects (i.e., verbal, numerical and analytical abilities) rather than the total ones. The mean value of efficiency and number of efficient schools of specification 3 are higher than those of specification 2. One likely explanation of this result is that the production model of specification 2 is a ‘subset’ of specification 3, which has more input/output variables. As more variables are included into the model, the greater is the chance that some inefficient schools will become efficient [26]. Having more variables in specification 3 will thus generally lead to higher average efficiency and more efficient schools, vis-à-vis specification 2.

To ascertain the efficiency results in “jackknifing” sensitivity approach, DEA results are tested to find the extent to which there were extreme outliers which affected the frontier, efficiency scores, and efficiency ranking in each of the DEA model specifications. For example, in the case of specification 1, each efficient school was dropped out one at a time in sequence, without replacement, and then the DEA model is re-estimated. The similarity of efficiency ranking between the DEA results with all schools included and those based on dropping out each efficient school one at a time was tested by the Spearman rank correlation coefficient. The mean efficiency from the iterations was also calculated. The results are summarized in Table 4.

TABLE 4
STABILITY OF DEA RESULTS TO OUTLIER SCORES, SPECIFICATIONS 1 – 3

Specification	No. of Efficient Schools	Range in Spearman Rank Correlation	
		Min	Max
1*	4	0.8548	0.9991
2	5	0.8770	0.9999
3	12	0.8947	1.0000

* Primal specification

The high rank correlation coefficients show that the rankings are relatively stable in regard to outlier schools determining the efficiency frontier. From the table, the variation of rank correlation coefficient was lowest in specification 1, ranging from 0.8548 to 0.9991. In other specifications, the ranking correlation was somewhat more stable, ranging from 0.8770 to 1.0000 in specification 2 and 3. As the outliers may affect to the stability of efficiency scores and ranking, e.g., due to the extreme difference of input-output combination, Specification 1 was somewhat more sensitive to outliers. Moreover, the greater the number of inputs and outputs in the additional DEA specifications, the greater the apparent stability of the DEA efficiency scores is in regard to outlier schools.

IV. CONCLUSION AND POLICY IMPLICATIONS

This study proposes the DEA technique to measure school efficiency, as DEA has been recognized as a practical method for this purpose. DEA uses benchmarking to measure efficiency of each school relative to others in its group. Such comparisons can assist in identifying efficient and inefficient schools within the group as well as indicating potential adjustment targets for the inefficient institutions. The empirical results reveal that the schools under observation, on average, operate inefficiently. To ascertain the robustness of the analytical framework and resulting measures, a sensitivity analysis was conducted by introducing two other specifications, different from the primal one. This was to examine whether modifications in the educational production function would result in substantive changes in the empirical findings on school efficiency. For each specification, the efficiency measures were re-estimated using the DEA model. Under jackknifing method, by dropping each outlier or efficient school one at a time in sequence without replacement and then the DEA model is re-estimated, the results suggest that efficiency ranking correlation was stable for all specifications. Moreover, under Spearman rank correlation, based on various additional school inputs and outputs, the results also indicate that the ranking between all specifications were almost similar.

This study, if validated, may have wide implications for national education policy. The performance-based DEA results indicate that the schools under observation, on average, operate inefficiently. Schools under observation have both efficient and inefficient schools and the adjustment targets of inefficient institutions, e.g., of SAT scores, indicate the extent to which these schools can improve their performance. If validated by further research, the findings could be used to prioritize institutions for treatment. In addition to allocating budget based on block grants or enrolments, policy-makers may use performance-based analysis to identify which institution will yield higher marginal returns, and which individual institution will give the most marginal returns.

For the future research direction, this study employs data only from public general upper secondary schools in Northern

Thailand; therefore, the principal limitation of this study is in the ability to generalize its findings to other level of educational institutions, or in other types of societies which is a limitation of studies in general. Moreover, this study assessed school performance through efficiency analysis only for a given period. Future research could focus on analyzing over a longer time or across subgroups or regions in a panel data setting. By doing so, a fuller understanding on the school efficiency analysis as well as the robustness of resulting measures on efficiency may be achieved. Further, this study uses inputs and outputs of the educational production function that may not be fully amenable to the other types of schools. Future studies might refine the inputs and outputs based on the objectives of each school type.

APPENDIX

DESCRIPTIVE STATISTICS OF INPUT, AND OUTPUT VARIABLES

Variables	Mean (S.D.)	Min.	Max.
Schools (n = 455)			
Inputs			
Teacher-student ratio (%)	4.77 (1.3945)	1.73	11.35
Proportion of students not from low-income families	0.52 (0.1966)	0.01	0.97
Teaching aide-student ratio (%)	0.60 (0.2828)	0.00	1.94
Outputs			
Passing	381.31 (381.3945)	9.00	2,389
Graduates	114.98 (119.9489)	2.00	779
National test score (SAT)	36.01 (5.5340)	25.45	67.81
SAT - Verbal Ability	13.33 (1.6837)	9.73	23.08
SAT - Numerical Ability	12.24 (2.2635)	8.00	26.00
SAT - Analytical Ability	10.44 (2.0142)	6.77	21.51

ACKNOWLEDGMENT

Author would like to thank Mr. Burin Kantabutra for the support to edit and share his time dedication and his crucial insights throughout this research works.

REFERENCE

[1] Y.C. Cheng, and T. *Townsend, Educational change and development in the Asia Pacific region: Trends and issues*. In T. Townsend & Y. C. Cheng (Eds.), *Education change and development in the Asia Pacific: Challenges for the future* (pp.318-343). Lisse, Netherlands: Swets & Zeitlinger, 2000.

[2] P. Hallinger, "Educational change in Southeast Asia: The challenge of creating learning systems." *J. of Ed. Admin.,* vol. 36, no. 5, pp. 492-509, 1998.

[3] A. Booth, "Education and economic development in Southeast Asia: Myths and realities." *ASEAN Econ. Bul.,* vol. 16, no. 3, pp. 290-306, 1999.

[4] OEC, *Education in Thailand 2004*, Bangkok: Amarin Printing and Publishing. 2004.

[5] B. Watcharasriroj, and J.C.S. Tang, "The effect of size and information technology on hospital efficiency." *J. of Hi. Tech. Manage. Res.*, vol. 15, no. 1, pp. 1-16, 2004.

[6] R. Manandhar, and J.C.S. Tang, "The evaluation of bank branch performance using data envelopment analysis: A framework." *J. of Hi. Tech. Manage. Res.*, vol.13, no. 1, pp. 1-17, 2002.

[7] B. Mante, and G. O'Brien, "Efficiency measurement of Australian public sector organizations: The case of state secondary schools in Victoria." *J. of Ed. Admin.*, vol. 40, no. 3, pp. 274-296, 2002.

[8] A. Charnes, , W.W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units." *Eur. J. of Oper. R.*, vol. 2, pp. 429-444, 1978.

[9] S. Bradley, G. Johnes, and J. Millington, "The effect of competition on the efficiency of secondary schools in England." *Eur. J. of Oper. R.*, vol. 135, no. 3, pp. 545-568, 2001.

[10] T. Kirjavainen, and H.A. Loikkanen, "Efficiency differences of Finnish senior secondary schools: An application of DEA and Tobit analysis." *Econ. of Ed. Rev.,* vol. 17, no. 4, pp. 377–394, 1998.

[11] M.J. Mancebón, and E. Bandrés, "Efficiency evaluation in secondary schools: The key role of model specification and of ex post analysis of results." *Ed. Econ.,* vol. 7, no. 2, pp. 131-152, 1999.

[12] A.C. Soteriou, E. Karahanna, C. Papanastasiou, and M.S. Diakourakis, "Using DEA to evaluate the efficiency of secondary schools: The case of Cyprus." *Inter. J. of Ed. Manage.*, vol. 12, no. 2, pp 65-73, 1998.

[13] S. Kantabutra, *Assessment of size effects on efficiency*, a proceeding paper for the 3rd International Conference on Engineering Education, Instructional Technology, Assessment, and E-learning (EIAE 07), Bridgeport, USA, 2007.

[14] S. Kantabutra, and J.C.S. Tang, "Urban-rural and size effects on school efficiency: The case of Northern Thailand." *Lead. and Pol. in Sch.,* vol. 5, no. 4, pp. 355-377, 2006.

[15] S. Kantabutra, and S. Kantabutra, *Investigating relationships between shared vision and public school performance: A proposed model*, a proceeding paper for International Conference on Making Educational Reform Happen: Learning from the Asian Experience and Comparative Perspectives, September 2004, Bangkok, Thailand, 2004.

[16] S. Kantabutra, and S. Kantabutra, *Heightening public school performance through vision-based leadership: A Thai perspective*, a proceeding paper for International Conference on Quality Educational Leadership: A Partnership of East and West, November 2005, Bangkok, Thailand, 2005.

[17] S. Kantabutra, and S. Kantabutra, *Relating shared vision components to Thai public school performance*, a proceeding paper for International Conference on Powerful Visions: Do they work in Asia, the 5th Annual Conference of the Asia Academy of Management, December 2006, Tokyo, Japan, 2006.

[18] F. Engert, "The reporting of school district efficiency: The adequacy of ratio measures." *J. of Public Budg., Acc. & Fin. Manage.*, vol. 8, no. 2, pp. 247-271, 1996.

[19] P.J. de Lancer, "Decision-making tools for public productivity improvement: A comparison of DEA to cost-benefit and regression analyses. *J. of Pub. Bud. Acc. and Fin. Ma.,* vol. 12, no. 4, pp. 625-646, 2000.

[20] D.F. Primont, and Domazlicky, B. "Student achievement and efficiency in Missouri schools and the No Child Left Behind Act.", *Econ. of Ed. Rev.,* vol. 25, no. 1, pp. 77-90, 2006.

[21] S. Waldo, "On the use of student data in efficiency analysis-Technical efficiency in Swedish upper secondary school." *Econ. of Ed. Rev.,* vol. 26, no. 2, pp. 173-185, 2007.

[22] T. Coelli, D.S.P Rao, and G.E. Battese, *An introduction to efficiency and productivity analysis*. Boston: Kluwer, 1998.

[23] P. Chalos, and J. Cherian, "An application of data envelopment analysis to public sector performance measurement and accountability." *J. of Acc. and Pub.Pol.*, vol. 14, pp. 143-160, 1995.

[24] W.J Fowler, Jr., and H.J. Walberg, "School size, characteristics, and outcomes." *Ed. Evalu. and Pol. Ana.,* vol. 13, no. 2, pp. 189-202, 1991.

[25] M.A. Muñiz, "Separating managerial inefficiency and external conditions in data envelopment analysis." *Eur. J. of Oper. R.*, vol. 143, pp. 625-643, 2002.

[26] J.A. Ganley, and J.S. Cubbin, *Public sector efficiency measurement. Applications of data envelopment analysis*. North-Holland, Amsterdam, 1992.

E-learning and Blended Learning in the Gulf Region

Dr.Soly Mathew Biju
College of Informatics and Computer Science
University of Wollongong in Dubai,
Dubai, United Arab Emirates.
SolyMathewBiju@uowdubai.ac.ae

Abstract

In today's ultra modern world the traditional face to face (F2F)teaching has been replace by e-learning or online learning .Companies find this method very effective to train their employees without using up work hours. E-learning has been implemented by many educational institutions at undergraduate and graduate level in the Gulf region. Blended learning has been adopted by many schools and universities in the Gulf. Blended learning is e-learning along with onsite support. This paper will discuss different modes of e-learning, assignment design techniques and E-learning trends in the Gulf region.

I. Introduction

There has been an increase in the number of universities and colleges implementing in E and blended learning as an integral part of instructional activities. They are trying to make best use of the technology available to promote e-learning and to motivate students to study outside the four walls of a classroom. These technological innovations have a direct impact on current university practices policies and have the potential of changing our traditional definition of education [1].E-learning can also mean extended learning that is learning outside the class room, from anywhere and at anytime.
Many schools in the gulf region are adopting this innovative style of teaching. Students find it interesting compared to the traditional method of teaching. Web-based instruction is already revolutionizing how students work, think , and access information [5].Some schools are using the free Learning Management Systems like 'Moodle ' which provides all the essential features required for blended learning.

Universities are also adoption tools like Blackboard.com and WebCT for online support along with the onsite teaching support. This can also be integrated with some classes being held in the second life which is 3D virtual world.

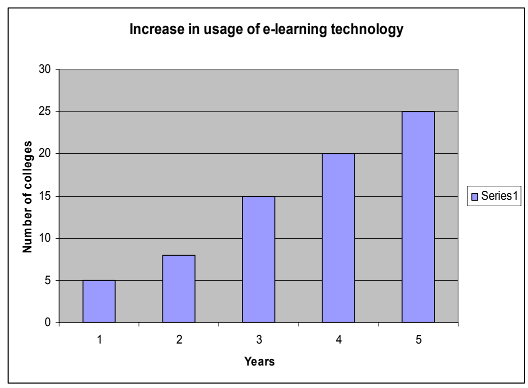


Figure1, increased usage of e-learning technology

As depicted in Figure 1, a recent survey conducted shows an increasing interest in e-learning over the years in the Gulf. Most of the universities in the gulf have a 'e-learning center for excellence'. It is very helpful at university levels for students who are working or who are frequently traveling to keep in touch with what is happening in their classes. It allows students who find it difficult to speak up in class to participate in online discussion forums. The students can participate in the online class discussion during their free time by logging on to the online class and posting his views to the discussion folder in an asynchronous mode.

The lecturers and teachers are able to save a lot of more time which can be utilized to develop and improvise on the course material.

E-learning can be very beneficial for students who are interested in continuing education along with their jobs and for people who have a lot of

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_2, © Springer Science+Business Media B.V. 2010

experience in the field or are currently working in the same field and want to enhance their knowledge about the subject. It does not require you to be physically present at a particular location. You can login to your class anytime from anywhere .This definitely saves a lot of time and helps the student to continue working while studying which is one of the major advantages. E-learning is a boon to specially those who have difficulty in travelling to and forth to attend the classes at the university.

II. Modes of online learning

A. Asynchronous mode

Online learning usually has two modes one is a completely asynchronous mode and synchronous mode. In case of asynchronous mode the lectures, assignment and examination questions are uploaded and available to the students online. The students can study from the lecture notes and specified text and then attempt the quizzes and question uploaded. There is no synchronised session between the lecturer and the student. Any query by the student will be send to the lecturer's mail id, who will then reply it within 24 hours in most of the online programs. Students used to the traditional mode of teaching may feel a little lost initially but will be able to cope up with online teaching.

B. Synchronous mode

In case of synchronous mode, the students and the instructors are required to login to the class at a particular time decided in advance. The lecturer will then conduct the class using PowerPoint presentation which will be visible to the student on his screen. There are many advance softwares that support this facility. The student will be able to listen to the explanation given for each of the slide displayed .He can ask a question to the instructor by clicking on an ion provided for the same which is very similar to raising your hand in the class to ask a question.
In this case you can have an experience of a virtual class room. The only drawback in this case is that the student should login to the class at a specific time given. This may be difficult and impossible for students who are constantly travelling. E-learning involves online class discussions where the students can login in their own free time and posting their opinions and views on the topic of discussion. The students are encourages to read each others suggestions and give their view about it. Data collected for this research shows that students enjoy the discussion session the most. It motivates them to login into the class more often.

III. Just-in-Time Teaching

"One Web-based approach, Just-in-Time Teaching (JiTT), was first devised by Novak, Patterson, Gavrin , and Christian (1999) to teach problem solving in physics" [3].This can be incorporated into online teaching .The students can be given assignments after every class and based on their performance in the assignment, the next lecture could be modified to suit the requirement of the class such that the it addresses the concepts the student have not understood correctly.
Universities in the Gulf region have adopted both these modes of e-learning. Many colleges in the U.A.E. prefer having a blended mode of teaching which is online teaching (asynchronous or synchronous) and onsite teaching(face to face) teaching.

IV. Designing Assignments

Assignments could be multiple choice questions or essay type questions. There are number of software vendors who provide standard softwares to support e-learning needs.. WebCT provides a facility wherein it is possible to create a question bank and then select the question for multiple choices. These questions banks are also available in the market for certain courses. These could be material that is designed for certain e-learning tools. eg. In WebCT, the instructor can select the required questions from the database of questions.
E-learning is not only converting text into an electronic form, it is making content more interactive and interesting with more audio and video effect.

The content design should be student centred. . From [4], a number of studies in recent years have highlighted critical aspects of learner readiness that need to be addressed. Some of these include:

1. Technology skills
2. Access to Technology
3. Technology Literacy and
4. Self-regulated learning

Points 1 to 3 can be addressed easily by the universities in the Gulf .Learning at self pace is a challenging task to achieve due to the fact that the courses are always designed for a particular period of time.

Online Group Projects:
It is very effective as the class can be divided into groups by the instructor and assigned different projects. Separate folder can created for each group. The student can use these folders to post their finding so that it is available to everyone within the group. They can also use the discussion forum or the chat session as a virtual meeting place. Instructor can keep a track of their participation and progress.

V. Other tools used for e-learning

E-learning classes could also be designed in Second Life. Second Life is a 3-D virtual world. There are virtual colleges and universities in the second life which have virtual classrooms and labs where experiments are carried out just like in a actual class room .This is very beneficial for some online programs .One such example could be a Chemistry class where experiences can be carried out in the virtual world. All the virtual chemicals used exhibits all the properties of the actual chemical. This also gives the student to explore and experiment without fear of any dangerous accidents. There a number of organizations in the Gulf region which provide training classes for flying a plane or driving a car is also provided online.
Colleges and universities are also using technologies like pod casting, wikis and blogs. Colleges, universities and schools in the Gulf are encouraging students to create e-portfolios which can then be given to potential employers.
An e-portfolio can be a web-based information management system that uses electronic media and services. The learner builds and maintains a digital repository of artifacts, which they can use to demonstrate competence and reflect on their learning. Having access to their records, digital repository, feedback and reflection students can achieve a greater understanding of their individual growth, career planning and CV building. Accreditation for prior and/or extra-curricular experiences and control over access makes the e-portfolio a powerful tool [2].

Mobile learning is a complete anywhere ,anytime learning technology.As shown in the figure it can acquire data and learning material not only from the internet but also from other mediums. Due to the successful development of Bluetooth, WAP (Wireless Application Protocol), GPRS (General Packet Radio System) and UMTS (Universal Mobile Telecommunications System), the technological structures for wireless telephony and wireless computing are now ready for use in the field of teaching. In most parts of the world wireless technologies and applications are replacing wired ones: e-Commerce is moving to m-Commerce; m-Business is replacing e-Business. The move to wireless technology in telephony and computing is irreversible.

In the education filed there has been a gradual change from traditional onsite learning to electronic learning (e-Learning) to mobile learning (m-Learning).

VI. Conclusion

The is a tremendous and fast increase in the number of colleges in the Gulf adopting for e-learning .E-learning is a very effective tool to making learning a very enjoyable and interesting experience. E-learning requires smart assignment design .The latest learning tool is the m-learning or mobile learning which is portable and a truly anywhere anytime technology. Further study needs to be conducted based on the technology used and favoured the most in the Gulf.

VII. References

[1]Daugherty M,Funke B.L. (1998) , 'University Faculty and Student Perceptions of Web-Based Instruction ',*Journal of Distance Education* ,1998 vol13.1
[2]E-portfolio.Available: http://www.danwilton.com/eportfolios/whatitis.php [Accessed 9th Jan 2008]
[3] Benedict ,James O.and Apple, Kevin J. 'E-xcellence in Teaching' *PsychTeacher Electronic Discussion List* July 2004.
[4]Oliver R. , Assuring The Quality Of Online Learning In Australian Higher Education. Paper presented at the "Moving Online Conference II" 2-4 September, Southern Cross University, GoldCoast ,Australia, Available: http://www.scu.edu.au/schools/sawd/moconf/MOC2_papers.html [Accessed 9th June 2007]
[5] Peck, K. L. and D. Dorricott, "Why Use Technology?" *Educational Leadership*, Vol. 51, No. 7, April 1994,pp. 11–14;

Scene Change Detection for Uncompressed Video

Patrick Seeling
Department of Computing and New Media Technologies
University of Wisconsin – Stevens Point
Stevens Point, Wisconsin 54481-3871
Email: pseeling@ieee.org

***Abstract*—In this paper, we present a scene change or shot boundary detection method based in the changes in entropy of differences between uncompressed video frames. As in the uncompressed domain, cues for scene or shot boundaries are not available, detecting video content features is a non-trivial and typically requires additional complexity in the evaluation. The entropy presents a metric for the complexity of information. Used on the differences between video frames, the entropy is able to measure the complexity of changes. We find that due to content dependency, however, the relative entropy changes in the sequence of video frames is a better indicator for detection.**

An evaluation of the presented approach finds that detection for a combination of video test sequences can be reliably performed using the U component of uncompressed YUV 4:2:2 video only.

I. Introduction and Related Work

The domain of video annotation, indexing, and retrieval has attracted a large body of research in the past. One of the main drivers for this area of research is the readily available computing power in normal desktop computers, digital video equipment, and motivation to share content over the Internet in various forms of access models, i.e., sharing via web pages, live streaming, and so on. It is expected that the amount of digital video that is available will continue to grow. With this vast amount of data available, annotation and logical indexing of video becomes a desired feature. One standard for the annotation of multimedia data is MPEG-7, see, e.g., [1]. Amongst different approaches to annotation, segmenting the video into its scenes or shots is the most intuitive and basic. Several different approaches exist to detect shot boundaries in video, see, e.g., [2], [3] for an extensive overview of different algorithms and their classifications.

Reliable and universal detection of scene changes or shot boundaries in the uncompressed domain in a fast manner, however, is still a challenge. While in the compressed video domain, typically cues can be obtained that are readily available from the encoded video stream (e.g., motion vector intensity or transform coefficients), these cues are not available in the uncompressed domain. To alleviate this problem, the authors in [4] employ a two-stage histogram-based method to determine scene changes and filter out unwanted false detections. More sophisticated approaches have to implement additional methods, such as edge detection, see, e.g., [5]. Other venues of research employ neural networks to determine scene changes based on several pixel- and color-based features of uncompressed video, see, e.g., [6]. In [7], the authors apply fuzzy logic approaches to detect scenes. Recently, the authors of [8] employed an autoregressive model based on the color histogram in the uncompressed domain to determine scene changes.

In the following, we present a scene detection algorithm that is solely based on the entropy of differences between frames. The motivation for using this approach is that with changes between frames, the entropy of the differences will increase. We evaluate the algorithm against a short video test sequence and a *Combined* video sequence, which is derived from multiple test sequences. We present the performance metrics precision and recall obtained with the proposed algorithm before we conclude.

II. Entropy of Frame Differences

The scene boundary detection method introduced in the following is applied on uncompressed video frames using the YUV 4:2:2 format. This format is typically used for video coding and transcoding for a variety of video codecs. The YUV format describes each individual pixel by its luminance (Y) and two color differences (U) and (V), also known as chrominance components.

We denote the i-th byte in the n-th frame out of $n = 0, 1, \ldots, N$ frames as $F_n(i)$. Furthermore, let $Y_n(i_Y)$ denote the luminance byte values and $U_n(i_{UV})$, $V_n(i_{UV})$ denote the i-th byte values for the two chrominance components. Note that due to the chroma sub-sampling, $0 \leq i_{UV} \leq \frac{i_Y}{4}$, i.e., the two chrominance components are restricted to half the resolution of the luminance component. Sub-sampling is used since the human eye is most sensitive to the luminance component. In the 4:2:2 format, the U and V values are each sub-sampled for a group of 4 luminance pixels and stored by component in a grouped manner on disk. The manner in which the individual values are sub-sampled and typically stored on disk is illustrated in Figure 1 for an individual frame.

We denote the probability for a specific byte value $F_n(i)$ in frame n as p_{F_n} with $F_n(i) \subset \{0, \ldots, 255\}, \forall i$. The entropy gives a measure for the complexity of an individual frame's information content. The entropy for the byte values of frame n is calculated as in Equation 1.

$$H_n = -\sum_{b=0}^{255} p_{F_n} \cdot \log_{256}(p_{F_n}) \tag{1}$$

The entropies for a frame's three individual components can be calculated in a similar manner.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_3,

In order to derive a better estimator for the changes between frames, we calculate the entropy for the differences between frames n and $n-1$ for all frames $n \geq 1$. The differences between two complete frames are calculated as in Equation 2.

$$F_{n,n-1}(i) = |F_n(i) - F_{n-1}(i)|_{\forall i} \quad (2)$$

The entropy for the frame differences is then calculated as in Equation 1 for $p_{F_{n,n-1}}$. The entropy of the difference frames (or the differences frames' components) can then be used to determine the scene boundaries by comparing subsequent difference frame entropies. We note that this approach requires at least three frames to be processed, as illustrated in Figure 2.

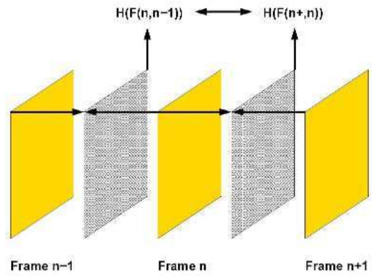


Fig. 2. Calculation of difference frames and entropy values.

Sequence	Frame	Content
News	0	A news sequence with two anchors and varying background.
	91	Background is one dancer close-up.
	151	Background is two dancers.
	241	Background is one dancer close-up.
Salesman	301	A salesman presents his product.
Akiyo	750	A news anchor talks in front of a static background.
Husky	1050	Several runners and their dogs, with some camera panning and zooming.
Bow	1300	A person enters, bows, and leaves.
Hall Monitor	1600	A hall monitor with people passing by.

TABLE I
DETAILS FOR THE CONCATENATED *Combined* VIDEO SEQUENCE.

III. SCENE CHANGE DETECTION METHOD

The initial evaluation of the algorithm presented in this paper is performed on the *News* sequence in the QCIF format ($i = 38016, i_Y = 25344, i_{UV} = 6336$). This sequence features two news anchors and a changing background with varying displays of dancers, as illustrated in Figure 3. Details for the *News* video sequence's content are provided in Table I. We illustrate the entropy for the frame differences in Figure 4. We observe that the entropy of frame differences "spikes" where the content of the underlying video sequence changes at frames 91, 262, and 241. This "spike" of the entropy can be used to detect the changes in scene content.

For the evaluation of the scene change algorithm introduced in Section III with respect to shot boundary detection, we now employ a *Combined* video sequence in the QCIF format, which is derived by concatenation of multiple video sequences. The sequences' details are given in Table I. The resulting entropy for the differences between complete frames $F_{n,n-1}$ is illustrated in Figure 5. We initially observe that the changes between individual scenes are represented by spikes in the entropy of the differences between full frames. We additionally observe that for each original sequence, a separate level and behavior of the entropy of frame differences can be observed in Figure 5. The scenes containing more motion and camera movement exhibit more varying entropies. For the *Husky* sequence, a generally high level of entropy is observed, while the *Bow* sequence's entropy exhibits more pronounced changes or "spikes" of the entropy.

The entropy of the differences for the individual frame components is illustrated in Figure 6. We observe that overall, all three individual components exhibit similar characteristics of the entropy compared to the complete frame. We note, however, that the two chrominance components exhibit a significantly lower level in their entropy values than the luminance component for the *Husky* and *Bow* sequences. We conclude that comparing the entropies of the difference frames for identification of scene boundaries itself is not advisable with respect to the different levels of entropy due to the content differences.

When the change in the entropies of the frame differences is evaluated, however, the scene changes can be detected

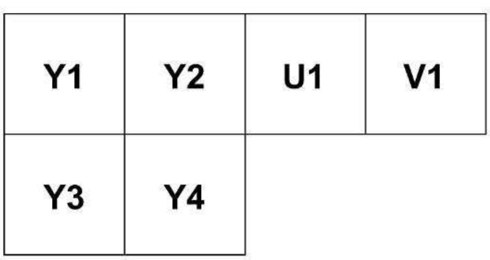


Fig. 1. YUV 4:2:2 pixel format and single video frame storage.

Fig. 3. Example screenshot of the *News* video sequence.

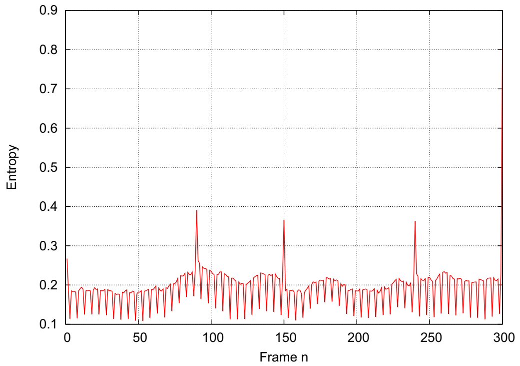


Fig. 4. Entropy of the frame differences for the *News* QCIF video sequence.

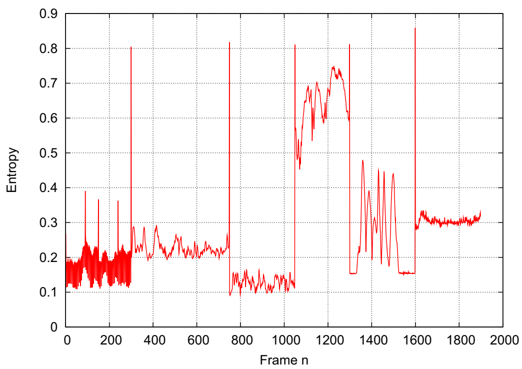


Fig. 5. Entropy of the frame differences for the *Combined* QCIF video sequence.

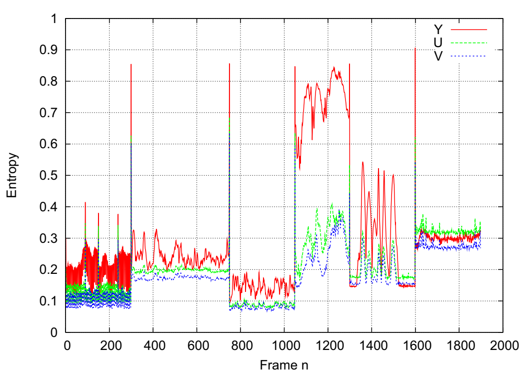


Fig. 6. Entropy of the frame differences for the individual Y,U, and V components of the *Combined* QCIF video sequence.

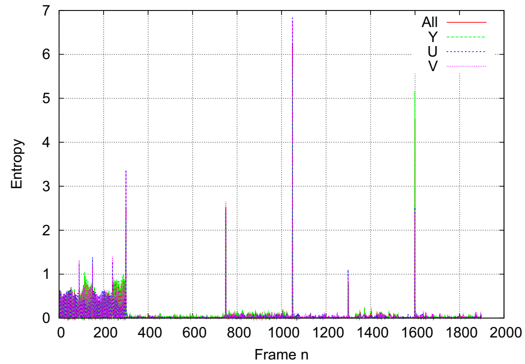


Fig. 7. Relative entropy changes for the frame differences of the *Combined* QCIF video sequence.

in a manner that takes the underlying content into account. As illustrated in Figures 5 and 6, the level of the entropy changes with the scene's content. The relative changes in the entropy values are hence a better estimator to determine scene boundaries. We calculate the relative entropy changes as

$$H_{n,n-1} = \frac{|H_n - H_{n-1}|}{H_{n-1}}. \tag{3}$$

The resulting relative entropy changes are illustrated for *Combined* video sequence in Figure 7. We initially observe that for the first part of the *Combined* video sequence, a fairly high level of noisy changes in the relative entropy can be examined, whereas this behavior is not visible for the other parts of the sequence. This behaviour can be explained by the background video in the *News* video sequence, see Figure 3. Thus, this behavior is content-dependent. Secondly, a closer examination of the "spikes" in the relative entropy yields that they occur at the scene change boundaries. We also note that these "spikes" are at least in the region of 100 percent changes in the relative entropy. This leads to the conclusion that changes of more than 100 percent in the relative entropy are indicators of scene changes.

To detect scene changes following this approach, we define the threshold for the detection as

$$H_{n,n-1} \geq t. \tag{4}$$

IV. Performance Evaluation

Typical performance metrics for the detection of shot and scene boundaries are *Recall* and *Precision*. The recall value presents a measure for the correct detection of changes, whereas the precision measures the correctness of detected changes. Let D denote the correct number of detections, D_F denote the number of false detections, and D_M denote the number of missed detections. Recall and precision can then

TABLE II
PRECISION AND RECALL VALUES FOR DIFFERENT THRESHOLDS FOR THE *Combined* VIDEO SEQUENCE.

Thres.	All		Y		U		V	
t	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
1	1	0.5	1	0.875	1	1	1	0.875
1.1	1	0.5	1	0.875	1	1	1	0.875
1.25	1	0.5	1	0.75	1	0.75	1	0.75
1.5	1	0.5	1	0.5	1	0.5	1	0.5
2	1	0.5	1	0.5	1	0.5	1	0.5

be calculated as

$$\text{Recall} = \frac{D}{D + D_M} \tag{5}$$

$$\text{Precision} = \frac{D}{D + D_F} \tag{6}$$

For the review of the proposed method on the *Combined* video sequence, we provide the values for precision and recall in Table IV for different thresholds t of the relative entropy. We observe that precision for the entropy of the complete frame differences is high for all components combined and individually. We furthermore note that there is no impact of different threshold levels on the precision. For the recall values, on the other hand, we observe a low value for the complete frame's entropy and slightly higher values for the Y and V components. Overall, only the U component yields complete detection of all scene changes with a threshold level close to 1.

An additional observation from this characteristic is the time required to parse the complete video, especially if resolutions above QCIF are considered. As the U component's size in bytes is only $\frac{1}{6}$ of the complete video frame's size, the scene detection speed can be greatly increased.

V. CONCLUSION AND OUTLOOK

In this paper, we presented a method to detect scene changes or shots based on the relative changes in the entropy of difference frames in the uncompressed domain. As the differences between frames increase, the entropy will increase as well. However, due to content dependency of the level and behavior of the entropy, we use the relative changes in the entropy of differences between frames for detection. We found that based on the performance metrics precision and recall, the fasted and most reliable approach to detect scene changes or shots is to calculate the relative entropy differences for the U component only.

Future research venues will include evaluation of the detection method presented herein for a larger variety of video content and additional refinements to detect additional changes in the underlying video based on the entropy of the frame differences.

REFERENCES

[1] P. Salembier and J. Smith, "Mpeg-7 multimedia description schemes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 748–759, Jun. 2001.

[2] R. Lienhardt, "Reliable transition detection in videos: A survey and practitioner's guide," *International Journal of Image and Graphics (IJIG)*, vol. 1, no. 3, pp. 469–486, 2001.

[3] C. Cotsaces, N. Nikolaidis, and I. Pitas, "Video shot detection and condensed representation," *IEEE Signal Processing Magazine*, vol. 23, no. 2, pp. 28–37, Mar. 2006.

[4] X. Yi and N. Ling, "Fast pixel-based video scene change detection," in *Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 4, Kobe, Japan, May 2005, pp. 3443–3446.

[5] R. Zabih, J. Miller, and K. Mai, "A feature-based algorithm for detecting and classification of production effects," *ACM Multimedia Systems*, vol. 7, no. 1, pp. 119–128, Jan. 1999.

[6] M.-H. Leea, H.-W. Yoob, and D.-S. Jang, "Video scene change detection using neural network: Improved art2," *Expert Systems with Applications*, vol. 31, no. 1, pp. 13–25, Jul. 2006.

[7] H. Fanga, J. Jiang, and Y. Feng, "A fuzzy logic approach for detection of video shot boundaries," *Pattern Recognition*, vol. 39, no. 11, pp. 2092–2100, Nov. 2006.

[8] W. Chen and Y.-J. Zhang, "Parametric model for video content analysis," *Pattern Recognition Letters*, vol. 29, no. 3, pp. 181–191, Feb. 2008.

ZigBee Performance in 400 KV Air Insulated Power Substation

Qingshan Shan[1], I A Glover[1], P J Moore[2], I E Portugues[2], R Rutherford[4], R Atkinson[3], Shahzad Bhatti[1], R J Watson[5]
[1]*Centre for Excellence in Signal & image Processing,*
[2]*Institute for Energy & Environment*
[3]*Centre for Intelligent Dynamic Communications*
Department of Electronic & Electrical Engineering, University of Strathclyde, 204 George Street, Glasgow G1 1XW, UK,
qingshan.shan@eee.strath.ac.uk
[4]*Scottish Power – Energy, Networks & Telecommunications, 1 Atlantic Quay, Glasgow G2 8SP, UK*
[5]*Department of Electronic & Electrical Engineering, University of Bath, Claverton Down, Bath BA2 7AY, UK*

***Abstract*-An experimental assessment of the impact of the electromagnetic environment of a 400 kV substation on the performance of ZigBee equipment is described. The experimental assessment includes a pragmatic field trial and a laboratory test. The laboratory test, in which all external noise and interference are excluded, is used as a control. A simple set of metrics are used to compare the performance of ZigBee equipment deployed in the substation with that deployed in a controlled laboratory environment. The results based on more than 1.6 Gbit of transmitted data show no significant adverse impact of the substation electromagnetic environment on the performance of ZigBee equipment**.

*Keywords***:** ZigBee, partial discharge, power substation, IEEE 802.15.4.

I. Introduction

The infrastructure investment in a national power transmission system is colossal. It is therefore necessary to operate such systems as efficiently as possible, consistent with maintaining acceptable security of supply. Efficient and reliable operation demands continuous monitoring of the system state resulting in instrumentation and control equipment being widely scattered throughout substation compounds. Information and control signals for both normal and abnormal operation are traditionally connected, using cables or optical fibres, to a SCADA (Supervision, Control and Data Acquisition) system [1] and/or its successor UCA (Utilities Communication Architecture) system [2]. Ethernet local area network (LAN) implementations of such UCA/SCADA systems, which simplify the addition/reconfiguration of instrumentation and the coordination of protection systems, have been proposed and are already being evaluated [3].

Significant flexibility and cost advantages over a wired LAN infrastructure would be gained, however, if signals could be routed around electricity substation compounds wirelessly. Furthermore, wireless communication technologies hold out the prospect of 'hot-line', sensors that can be deployed on energized high-voltage (HV) equipment without the inconvenience and costs associated with bridging the system's primary insulation [4, 5]. WLAN and WPAN technologies represent obvious opportunities to realize these advantages.

The casual deployment of wireless technologies for critical functions is not, however, without risk. Whilst the naturally occurring noise environment is relatively benign at WLAN and WPAN frequencies [6] the man-made noise environment within a substation compound is complex and hostile due, for example, to PD from imperfect insulation and sferic radiation from switching and fault transients. (The term sferic usually relates to radiation from a lightening event but is used here as a shorthand for similar radiation arising from any large current transient.) The latter is of particular concern since it is on just such occasions that control and protection equipment is required to operate reliably. It is possibly for this reason that UCA demonstration systems have until now employed a 'wired' (often fibre) transmission medium e.g. [7].

An investigation into the vulnerability of WLAN and WPAN technologies to impulsive noise in electricity transmission substations has been proposed [8]. One of the project objectives requires an assessment of ZigBee technology in impulsive noise environments and its suitability for deployment in HV substations. As part of this assessment a laboratory test and field trial of ZigBee equipment have been carried out.

The field trial comprises a deployment of ZigBee equipment in a 400 kV electricity supply substation. The laboratory test replaces the antennas and radio path between Zigbee transceivers with high-quality coaxial cable and appropriate microwave attenuators.

This paper describes partial discharge as observed in electricity supply substations, briefly reviews the relevant aspects of ZigBee technology, and presents the results of both laboratory test and field trial.

II. PD in Substations

An electrical discharge is partial if it fails to fully bridge the space between a pair of electrodes. It can occur around an electrode in a gas (corona), within gas bubbles in a liquid or within the space created by voids in a solid. HV plant (transformers, switchgear, cables etc) is especially prone to PD

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_4,

if its insulation is damaged and/or as its insulation ages. If remedial action is not taken the insulation can be seriously compromised leading, ultimately, to catastrophic failure. PD current pulses in strong insulators (e.g. SF_6) can have rise-times as short as 50 ps and may contain significant energy at frequencies up to 3 GHz [9]. Such impulsive signals can give rise to electromagnetic resonances within the conducting enclosures in which they occur.

PD propagation within a Gas Insulated Substation (GIS) is by a combination of transverse electric (TE), transverse magnetic (TM) and transverse electromagnetic (TEM) modes [10]. Laboratory tests have suggested that two principal mechanisms are responsible for PD signal damping. These are reflections due to characteristic impedance discontinuities (such as is caused by spacers) and energy conversion from TEM to TE or TM modes [11]. Although the character of PD appears to have some dependence on the size and geometry of plant components (e.g. insulating spacers, L-shaped buses, T-branch buses) damping typically appears to become significant somewhere between 100 MHz and 300 MHz and increases with increasing frequency above this [12]. PD energy in the frequency range 0.5 - 1.2 GHz, however, is readily radiated from apertures formed, for example, by insulating spacers or bushings [13].

Energy from PD processes can be radiated whenever spectral components arising from current pulse edges extend into the radio frequency (RF) region [14]. Signals radiated from open-air substations are typically stronger than those from underground substations due, in the latter case, to an enclosing metallic tank located in a steel-reinforced concrete building [15].

III. ZigBee

ZigBee is a specification of a network and security services application layer technology developed by the ZigBee Alliance and based on the IEEE 802.15.4 standard. It is a low-cost, low-power, two-way wireless communications technology. It is primarily intended for domestic and commercial applications e.g. home and building automation, PC peripherals, toys and games but also finds application in other contexts, e.g. medical sensors and industrial control [16]. The IEEE 802.15.4 standard addresses the physical (PHY) and media access control (MAC) layers and defines a 250 kbit/s, 32-chip, direct sequence spread spectrum (DSSS), offset quadrature phase shift keying (OQPSK) radio signal operating in the 2.4 GHz unlicensed Industrial, Scientific and Medical (ISM) band. Alternatives bands exist around 868 MHz and 915 MHz. From a total of 27 frequency channels, one lies in the 868 MHz band, ten in the 915 MHz band and 16 in the 2.4 GHz band. In addition to providing link-layer security including access control, confidentiality, message integrity and optional message freshness, it employs network- and application-layer security services, a 128-bit link key to secure pair-wise communications and a 128-bit network key to secure broadcast communications [17]. A commercial system-on-chip implementation of a 2.4 GHz IEEE 802.15.4 compliant transceiver is already available [18, 19].

Previous work on the impact of interference on ZigBee systems has concentrated on the coexistence of this technology with similar technologies, especially WLAN and other WPAN transceivers. ZigBee has been shown to be potentially vulnerable to IEEE 802.11b (WiFi) and IEEE 802.15.1 (Bluetooth) transmissions. An analytical model suggests, however, that the IEEE 802.15.4 network should have little effect on the performance of IEEE 802.11b networks [20]. It has been concluded, from practical measurements, that IEEE 802.15.4 transmissions produce no significant impairments on the operation of IEEE 802.11 equipment, but that the converse is not necessarily true, i.e. IEEE 802.11 transmissions might degrade the performance of an IEEE 802.15.4 network [21]. The measurements reported appear to show that interference from IEEE 802.11b transmissions does not significantly impact IEEE 802.15.4 performance providing a physical separation of at least 8 m is maintained [22]. (References [21] and [22] suggest that a frequency offset of at least 7 MHz in operational frequencies is required for a satisfactory IEEE 802.15.4 performance.)

The authors are not aware of any similar work to establish the interfering effects on ZigBee transceivers of impulsive noise of PD origin as found in electricity supply substations. The practical field trial and laboratory test described in section 4 address this.

IV. Field Trial and Laboratory Test

The field trial was carried out at Strathaven 400 kV Air Insulated Substation (AIS). The trial assesses the performance of ZigBee in this challenging electromagnetic environment. The laboratory test was carried out in the Geoffrey Smith Intelligent Dynamic Communications Laboratory at the University of Strathclyde. The laboratory test represents a control, i.e. a test which replicates that of the field trial but excludes all external noise and interference. The methodologies are described below.

A. Field Trial

The field trial system consists of two terminals, one a data source and the other a data sink, as shown in Figure 1. The data source and data sink terminals comprise a ZigBee module interfaced to a laptop computer. The ZigBee modules are based on the Ember EM250 ZigBeeTM/IEEE802.15.4 chip. The RF

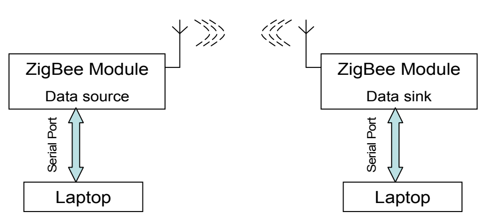


Fig.1. Field trial system

frequency band is 2.4 GHz and a total of 16 channels (IEEE 802.15.4 channels 11 to 26) are used. The maximum air-interface data throughput is 250 kbit/s. The maximum RF output power is 4 dBm. The receiver sensitivity with nominal 1% packet error rate is -98 dBm. Communication between the ZigBee module and the 2.33 GHz, 1 GB RAM, 120 GB HDD laptop computer is via an RS232 serial port.

The data source and data sink were located in different rooms of the 400 kV control building within the substation compound. The transmitter powers of the ZigBee modules were set to their maximum of 5 dBm. The separation between transmitter and receiver terminals was 21 m. The data sink signal strength indicator (RSSI) was -87 dBm which was 11 dB above the ZigBee receiver sensitivity. 21 m therefore represents close to the maximum separation in this environment. The serial port settings for interface between ZigBee modules and laptop computers are 115.2 kbit/s, 8 data bits, 1 stop bit, 1 parity bit (even) and hardware control flow.

Data was transmitted constantly from the source to the sink. Two transmission modes have been implemented; raw data transmission (mode 1) and data transmission with cyclic redundancy check (mode 2). In the former scheme data is neither encoded nor error checked. In the latter scheme data are CRC (cyclic redundancy check) encoded at transmitter and checked for errors at receiver.

All data transmissions are of pseudo fixed code blocks. The block lengths are 114 symbols for mode 1 transmissions and 65 symbols for mode 2 transmissions. These are the maximum block lengths specified by the ZigBee chip manufacturer.

The durations of the trials were a little more than four days and 10 days for mode 1 and mode 2 trials, respectively. These durations were chosen to yield a comparable volume of transferred data in both cases.

B. Laboratory Test

For the laboratory test terminal hardware, interface and settings were identical to those in the field trial. The communications channel between data source and sink, however, was replaced with a microwave cable and adjustable microwave attenuators, as shown in Figure 2. The cable and attenuators are specified for operation between DC and 18 GHz, and DC and 20 GHz, respectively. The ZigBee modules are enclosed in metallic boxes to provide shielding from external electromagnetic interference.

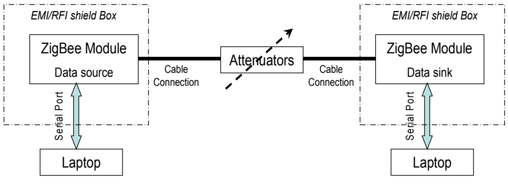


Fig. 2. Laboratory test system

The attenuators between source and sink modules were adjusted to give an RSSI of -84 dBm for mode 1 transmission and an RSSI of -87 dBm for mode 2 transmission. -84 dBm was found to be required for mode 1 transmissions (i.e. 3 dB more than that used for mode 2) in order to avoid complete system failure. This received power level therefore represented the minimum required to avoid the risk of transmission termination.

V. RESULTS AND DISCUSSION

Data transmission errors can be caused by external noise and interferences or internal noise (including clock jitter). Two types of error can be distinguished: (i) symbol (byte) error and, (ii) symbol (byte) loss. Symbol error refers to a received symbol being different from that transmitted and is directly linked to bit errors. Symbol loss refers to symbols transmitted from the data source that remain detected at the data sink. Bit error ratio (BER), symbol error ratio (SER), symbol loss ratio (SLR) and overall symbol error and loss ratio (OSER) are used here as performance metrics.

ZigBee performance for mode 1 transmission is summarized in Table 1 for both field trial and laboratory test. Performance for mode 2 transmission is summarized in Table 2.

Table I shows that the mode 1 laboratory test SER and BER are better than those of the field trial by factors of 1.45 and 1.48 respectively. Laboratory test SLR and OSER, however, are worse than those for the field trial by factors of 3.65 and 3.59 respectively.

Table II shows that the mode 2 field trial SER, SLR, OSER and BER are worse than those of laboratory test by factors of 1.03, 1.60, 1.54 and 1.03, respectively.

These results taken together suggest that (i) symbol loss dominates symbol error and (ii) the electromagnetic environment of the substation did not significantly affect ZigBee performance.

For mode 1, symbol loss probability was greater for the ideal (laboratory) channel than in the substation trial, despite RSSI being 3 dB greater. It is thought that this may be explained by RF signal leakage from the antenna connector or PCB connector track arising due to the high attenuation path (> 73 dB) between data source and sink. Such leakage present in the electrically screened enclosure might be reflected back into, or otherwise recaptured by, the transmitter, thereby disturbing the module transceiver.

TABLE I

PERFORMANCE SUMMARY FOR MODE 1 (RAW DATA) TRANSMISSION

	Field Trial	Laboratory Test
RSSI	-87 dBm	-84 dBm
Number of transmitted symbols	204,825,738	204,825,852
Symbol error ratio (SER)	1.9578×10^{-6}	1.3475×10^{-6}
Symbol loss ratio (SLR)	9.5691×10^{-5}	3.4928×10^{-4}
Overall symbol error & loss ratio (OSER)	9.7649×10^{-5}	3.5062×10^{-4}
Number of transmitted bits	1,638,605,904	1,638,606,816
Bit error ratio (BER)	5.2179×10^{-7}	3.5335×10^{-7}

TABLE II

PERFORMANCE SUMMARY FOR MODE 2 (CRC PROTECTED) TRANSMISSION

	Field trials	Laboratory Test
RSSI	-87 dBm	-87 dBm
Number of transmitted symbols	230,806,745	230,806,810
Symbol error ratio (SER)	2.7945×10^{-6}	2.7036×10^{-6}
Symbol loss ratio (SLR)	3.446×10^{-5}	2.1516×10^{-5}
Overall symbol error & loss ratio (OSER)	3.7252×10^{-5}	2.4219×10^{-5}
Number of transmitted bits	1,846,453,960	1,846,454,480
Bit error ratio (BER)	7.4196×10^{-7}	7.1813×10^{-7}

VI. CONCLUSION

A field trial to determine the practical performance of ZigBee technology operating in the severe electromagnetic environment of a 400 kV electricity supply substation has been described. A laboratory test which excludes all external noise and interference and which has been conducted as a control has also been reported. The results of trial and test suggest that there was no significant adverse impact on the performance of ZigBee technology by the electromagnetic environment (including partial discharge) of the substation.

ACKNOWLEDGMENT

The UK Engineering & Physical Sciences Research Council (EPSRC) is gratefully acknowledged for her financial support of this project.

REFERENCES

[1] IEEE," Definition, Specification, and Analysis of Systems Used for Supervisory Control, Data Acquisition, and Automatic Control," in *IEEE Standard C37.1*, 1994.

[2] IEC " Communications Networks and Systems in Substations-Part1: Introduction and overview," in *IEC Standard 61850-1*, 2003.

[3] R. Ehlers, "Using UCA to integrate and deliver diagnostic data from substation," in *12th substation equipment diagnostics conference* New Orleans, USA, 2004.

[4] H. Sedding, S. Bomben, S. Rizzeto, and R. Densley, "Assessment of station cable condition using wireless telemetry and diagnostic tests," in *2001 EEIC conference*, 2001.

[5] P. Duggan, P. DilLilo, P. DiScala, R. Schwabe, J. Braun, N. Fujimoto, and S. Rizzetto, "Field experience with live wireless telemetry sensors on power equipment operating to 345kV," in *12th substation equipment diagnostics conference* New Orleans, USA, 2004.

[6] "Radio noise," in *ITU-R Recommendation*, 2003, pp. 372-378.

[7] B. Smith, C. McClure, and R. Ehlers, "Integrating UCA at the TVA/Tiptonville switching station for SCADA and relay protection applications," in *12th substation equipment diagnostics conference* New Orleans, USA, 2004.

[8] I. A. Glover, Q. Shan, P. J. Mooore, I. E. Portugues, and R. J. Watson "An investigation into the vulnerability of WLAN and WPAN technologies to impulsive noise in electricity transmission substations," in *Ninth International Symposium on Communication Theory and Applications ISCTA'07*, Ambleside, Lake District, UK, 2007.

[9] M. D. Judd, O. Farish, J. S. Pearson, and B. F. Hampton, "Dielectric windows for UHF partial discharge detection," *Dielectrics and Electrical Insulation, IEEE Transactions on [see also Electrical Insulation, IEEE Transactions on],* vol. 8, pp. 953-958, 2001.

[10] C. Chang, C. S. Chang, J. Jin, T. Hoshino, M. Hanai, and N. Kobayashi, "Source classification of partial discharge for gas insulated substation using waveshape pattern recognition," *Dielectrics and Electrical Insulation, IEEE Transactions on [see also Electrical Insulation, IEEE Transactions on],* vol. 12, pp. 374-386, 2005.

[11] H. Imagawa, K. Emoto, H. Murase, H. Koyama, R. Tsuge, S. Maruyama, and T. Sakakibara, "PD signal propagation characteristics in GIS and its location system by frequency components comparison," *Power Delivery, IEEE Transactions on,* vol. 16, pp. 564-570, 2001.

[12] T. Sakakibara, H. Murase, E. Haginomori, S. Wakabayashi, K. Emoto, and A. Ogawa, "Study of propagation phenomena of partial discharge pulses in gas insulated substation," *Power Delivery, IEEE Transactions on,* vol. 13, pp. 768-776, 1998.

[13] T. Hoshino, K. Kato, N. Hayakawa, and H. Okuno, "Frequency characteristics of electromagnetic wave radiated from GIS apertures," *Power Delivery, IEEE Transactions on,* vol. 16, pp. 552-557, 2001.

[14] P. J. Moore, I. E. Portugues, and I. A. Glover, "Radiometric location of partial discharge sources on energized high-Voltage plant," *Power Delivery, IEEE Transactions on,* vol. 20, pp. 2264-2272, 2005.

[15] M. Hikita, H. Yamashita, T. Hoshino, T. Kato, N. Hayakawa, T. Ueda, and H. Okubo, "Electromagnetic noise spectrum caused by partial discharge in air at high voltage substations," *Power Delivery, IEEE Transactions on,* vol. 13, pp. 434-439, 1998.

[16] Z. Alliance, "ZigBee Specification," 1 December 2006, 2006.

[17] Z. Jianliang, M. J. Lee, and M. Anshel, "Toward Secure Low Rate Wireless Personal Area Networks," *Mobile Computing, IEEE Transactions on,* vol. 5, pp. 1361-1373, 2006.

[18] W. Andrew, "Commercial Applications of Wireless Sensor Networks Using ZigBee," *Communications Magazine, IEEE,* vol. 45, pp. 70-77, 2007.

[19] W. Kluge, F. Poegel, H. Roller, M. Lange, T. Ferchland, L. Dathe, and D. Eggert, "A Fully Integrated 2.4-GHz IEEE 802.15.4-Compliant Transceiver for ZigBee™ Applications," *Solid-State Circuits, IEEE Journal of,* vol. 41, pp. 2767-2775, 2006.

[20] I. Howitt and J. A. Gutierrez, "IEEE 802.15.4 low rate - wireless personal area network coexistence issues," in *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE* 2003, pp. 1481 - 1486.

[21] M. Petrova, J. Riihijarvi, P. Mahonen, and S. Labella, "Performance Study of IEEE 802.15.4 Using Measurements and Simulations," 2006.

[22] D. G. Yoon, S. Y. Shin, W. H. Kwon, and H. S. Park, "Packet Error Rate Analysis of IEEE 802.11b under IEEE 802.15.4 Interference," in *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd*, 2006,pp.1186-1190.

Content Management System Effort Estimation Using Bagging Predictors

Naveen Aggarwal*, Nupur Prakash**, Sanjeev Sofat***
* CSE Deptt, University Instt. Of Engg. & Technology, Panjab University, Chandigarh, INDIA
**School of Information Technology, GGSIP University, Delhi, INDIA
*** CSE Deptt. Punjab Engg. College, Chandigarh, INDIA
e-mail: navagg@gmail.com

Abstract - This paper presents an effort estimation model for the content management systems which can be used to estimate the effort required for designing and developing. The data from the different content management system projects are studied and the model is finalized by using the bagging predictor on linear regression learning model. These projects are categorized based on their size and total/build effort ratio. The size of the project is estimated by using the modified object point analysis approach. A questionnaire is prepared to help project managers to find out the different objects, their categories and their complexity in the project. Final effort is estimated using the project size and the different adjustment factors. For better calculation of these adjustments factors, these are categorized based on their characteristics viz. Production and General system characteristics, Developer's experience and capability. Another questionnaire is used to refine the model and it has to be filled by the project managers after completing the project. The proposed model is validated by studying twelve completed projects taken from industry and seventy different projects completed by the students. The proposed model shows a great improvement as compared to the earlier models used in effort estimation of content management system projects.

I. INTRODUCTION

The way the information is currently being accessed on web may get change drastically in the coming years with the Web 2.0. The creativity of user is enhanced as numerous ways can be devised to collaborate. But with this, there is a challenge of efficient web design that gives independence to share information in a user defined way [1]. The Content Management System (CMS) tools have already started playing this role and their use in developing and maintaining the web sites is increasing day by day. This is mainly because of their ability to enable the companies to build, deploy and maintain highly dynamic internet, intranet and extranet web sites, quickly and efficiently. The processing engine of CMS dynamically creates hypermedia pages using the different multimedia contents, which are stored as objects on different server [2, 3]. These objects are stored on the smart server cache and served to the different devices and individuals. The effort required to design such hypermedia application is very high because of the high coupling among the elements. Further designing customizable screens which user can change for better information sharing is always difficult than normal screens. In a typical business environment with changing customer needs, the designs are always inconsistent. This may lead to underestimation or overestimation of the resources [4]. The underestimation causes understaffing, under scoping of QA effort, and mismanagement of schedule. Similarly overstaffing is the main cause for cost increments, time increments and overuse of resources. These effects have large impact on other concurrently going on projects. Further as software grew in size and importance, it also grew in its complexity, making it very difficult to accurately predict the cost of the software development. The software development process of web CMS requires gradual refinement. Size estimation in these projects is very difficult because of multimedia contents having different characteristics, access methods and storage requirements [5]. So it is hard to find basis for good estimates in web CMS. It is observed that the better accuracy in estimation can be achieved by developing the domain based models.

The main aim of this research is to propose the effort estimation model for the content management systems. For the size estimation of the project, the different techniques such as function point analysis, object point analysis and use case analysis are compared. An Estimation model, to address the estimation needs for the client in the projects that uses tools like CMS, ITG etc, is developed using data from the Subject Company and different student projects executed at Panjab University, Chandigarh. The estimation model is finalized by using the bagging predictors.

The rest of the paper is organized as follows. Section 2 reviews the different effort estimation model for web based projects. The proposed effort estimation model is discussed in detail in section 3. Section 4 discusses the validity of the model and represents the results. Finally, Section 5 concludes the paper.

II. RELATED WORK

Over the years, different approaches have been suggested for the estimation of different types of projects. These projects are categorized as development projects; maintenance projects and support services by keeping in mind that these activities never occurred simultaneously. Most of these approaches estimate the effort based on the size of the development project and same approach is also used in estimating the effort of web projects [6].

Most of the web development companies are very small and are working at SEI CMM level 2. Hence a study was

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_5,

conducted by Ruhe et al [7] to assess whether the COBRA (Cost Estimation Benchmarking and Risk Analysis) method was adequate for estimating Web development effort accurately using data from these companies. As COBRA is heavily depended upon the expert opinion and data on past projects to estimate development effort hence accurate results are possible only if size estimates are correct. Although the prediction accuracy obtained using COBRA for similar project is very good but for new projects some learning oriented method need to be incorporated.

Mendes et al. [8] studied the different learning oriented techniques such as the case-based reasoning, linear regression, and stepwise regression techniques with an aim to estimate development effort for Web applications. He concluded that the case based reasoning gives the best result for experienced student datasets. In 2001, Mendes et al. [9] studied the prediction accuracy of these approaches using linear and stepwise multiple regression models. This time data was collected from the web sites of experienced programmers by considering the three main entities such as application page, media and program. In each entity the following five factors are considered such as: length size, reusability, complexity size, effort, and confounding factors. Results showed that the best predictions were obtained for the entity program, based on non-reused program measures (code length and code comment length). In this study, results were not very good for application page entity. So in the next study, Fewster et al [10] proposed a generalized linear model for estimation of effort. This model provides a flexible regression framework for predictive modeling of effort. But the model was very much dependent on the accuracy of the different data and size metrics used.

Reifer et al [11] proposed an extension to the COCOMO-II model by introducing WEBMO model. Different cost drivers were proposed by keeping in mind the demands of web projects. Further, as web projects keep on changing, hence variable effort power law was used rather than the fixed one. In WEBMO, size was measured by using analytical Halstead’s formula for volume. Although, all the different sub components in a typical web project were considered, but with the changing development tools, the fixed cost drivers are not suitable. Morisio et al [12] solved this problem by introducing the estimation model for the web development companies to predict and track the development costs when new technologies are employed. An object oriented framework was used to gather the data and a learning factor was associated which changes with every usage of framework. The results clearly show that effort to write a code reduces with every usage of the framework.

In 2003, Baresi et al. [13] conducts a study with an aim to estimate the design effort of web applications. Their dataset comprised 30 Web applications developed by students. The relationship between the numbers of size measures were obtained and organized in categories. The categories employed were information model, navigation model and presentation model. They identified a few attributes that may be related to the total design effort. In addition, they also carried out a finer grain analysis, to find out which of the used measures have an impact on the design effort when using W2000 [14]. But all of these approaches discussed above are not directly applicable to the web content management system projects. This is mainly because:

1. There are no well defined guidelines to identify the development model for CMS. Most of the processes employed are ad hoc.
2. There are no standard tools and technology defined for CMS. It is very difficult to estimate the size of the project when the developers are not experienced. No mechanism available to create estimations model for new technologies and tools.
3. Existing web project estimation guidelines are not robust enough to take care of the complexities of the interfacing system in CMS.

Braga et al [15] suggested that the bagging predictors [16] can be used to improve the performance of any methodology that uses linear regression, regression trees or support vector regression for estimation. But still, it is felt that for different technologies which are being used in different domain, there should be different estimation model which are developed just from the experience gained in that domain only.

III. EFFORT ESTIMATION MODEL

The basic approach of the Content Management System Effort Estimation Model (CMSEEM) is described in the Figure 1. The initial effort estimation model is an extension of the COCOMO 2.0 Model [17]. The data collected from the twelve different content management system projects taken from the subject company. It is assumed that all the data of twelve different project requests studied is correct without any deviation. The model is then refined from the data collected from the 70 different CMS projects completed by the post graduate students. All the students who are involved in developing projects are provided with a detailed set of project requirements. After studying the requirements, they need to certify the following points before they start developing the projects.

- Requirements Specifications are clear and well understood
- Tools used to develop the project can be mapped to Object Point analysis [18] i.e. from Requirement analysis, it is possible to count the total number of objects required for development in the specified tool.
- Overall size of project is small and do not involve any research activity.

All the students were given a questionnaire, which they have to submit after completing the software requirement specification document. The questionnaire is about the 40 different parameters related to the setup, implementation, design, performance, user interface, usability, reliability, external interface, and other requirements of the project.

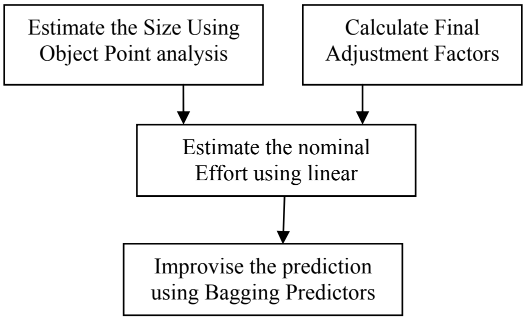


Figure 1: Approach for Effort Estimation of CMS

A. Estimation of Size

For Size calculation, we have opted for the Object Point Analysis. Object Points are a measure of the size of computer applications and the projects that build them. The size is measured from a components point of view. It is independent of the computer language, development methodology, technology or capability of the project team used to develop the application. Following steps are mainly followed to calculate the total number of Object Points of a particular project.

Step 1: Assess Object Counts: All the different categories of possible components in CMS are recognized [19-20]. These components and categories may vary according to the tools used.

Step 2: Each object instance is classified into corresponding total number of object points depending upon the value of its characteristic dimensions. This value depends upon the different characteristic of the objects which may have additive or multiplicative effect on it. Brief description about the different object instances categories and their characteristics is as follows.

- Branch Setup: In Branch Setup, Main Multiplicative factor is total number of branches and sub branches. Other tasks such as setting up work area, creating template data, changing configuration settings (setting proxies, receivers, color palettes and visual formats) are counted as additive task.
- Authorities Setup: Number of functionalities to be enhanced in a CMS is taken as multiplicative factor. The number of tasks need to be performed for each of the functionality are taken as additive tasks such as changing permissions, events and privileges for each users, security settings etc.
- Work Flow Setup: Different tasks (user tasks, group tasks, local, external, CGI tasks, update, submit, end tasks etc) and different directives (Team Info, CGI Info, element Info etc) in work flow are taken as multiplicative factors. For each task different dependencies are taken as additive factors.
- Static Containers: Total number of different items in each container is taken as multiplicative factor. For each of these items, number of different characteristics and properties settings and other GUI settings are considered as an additive factor.
- Dynamic Containers: No of Object Points for Dynamic Containers is counted as sum of object points for the extra Template created and object points of container by considering it as static with current data items.
- Data Components: For each data component, the total number of validations required for that component is considered as multiplicative factor. For each of these data components, number of source of data tables from where it is taking its data or sending its data along with the different hyperlinks associated with it are considered as additive factor for that item.
- External Interface Components: For this category, the total number of internal objects interlinked with external objects is taken as multiplicative factor. For each internal object, total number of external objects, API, ODBC, and tools.are taken as an additive factor.
- Screens: Different data forms, message boxes, error messages, tabs, menus, frames etc are considered as different views. Total number of such views is considered as multiplicative factor. For each view, its interaction with databases, validations using Java Script functions, operations and different style classes used are taken as additive factors.
- Reports / Files: Different sections in each report / file are considered as multiplicative factors. In each section, the number and source of data are considered as the additive factors.
- Database Setup: Different tables/views or triggers in particular database are considered as multiplicative factors. In each table/view/trigger, the different attributes and any changes made to these attributes are taken as additive factors.

For all the objects, their multiplicative and additive factors are used to calculate the final object point count.

Step 3: Estimate the expected percentage of reuse to be achieved for the different Object Categories in this project. Compute the New Object Points (NOP) to be developed,

$$NOP= (Object\ Points)\ X\ (100- \%\ Reuse) / 100 \qquad (1)$$

Step 4: Determine the Object Points: Add all the weighted object instances to get one number, the object point count.

B. Estimation of Adjustment Factors

The adjustment factors affect the feature overall and not any object in particular. Main three adjustment factors considered are:

- Production System Characteristics (PSC)
- General System Characteristics (GSC)
- Developer's Experience and Capability (DEC) [21]

Main Production System Characteristics (PSC) that affects the system can be categorized as follows:

- Application Characteristics (Round the Clock Support, Criticality, Complexity)

- Data Characteristics (Volume, Quality, Change Dynamics, External Dependency)
- Others (Security, Portability, Privacy, Audit ability, User training etc.)

The weighted influence of each of these characteristics is added to get the Adjustment factor corresponding to the PSC. If a characteristic has strong influence then 3 points are added, for average influence, 2 points are added and for less influence, one point is added.

The General system characteristics such as effective GUI, distributed data processing, installation and operational ease also have the direct impact on the effort. The complexity of GUI increases with the amount of control provided to user. The other factors that increase it are consistent interface with visual, conceptual and linguistic clarity and support to different languages. In most of the CMS, data is processed dynamically using the appropriate component of the system. These components may be located at different servers used by the CMS. The total number of such components in a particular system defines the complexity of that system. Further the total different types of servers used to store data and different types of network used to connect these, increases the complexity.

Some of the CMSs, require automated installation and operation without any operator intervention. Developing such applications is very complex. Similarly, the applications designed for automated recovery from failure or errors have more degree of complexity. The effects of all such general system characteristics are calculated on the same lines as we have calculated the PSC.

Finally, the effect of developer's experience and capability on the estimated effort can be considered in the similar way as suggested by Boehm et al [21].

The Final adjustment factor is calculated as follows

Final Adjustment factor (AF) =

((Sum of PSCs & GSCs) * 1/100) + 0.5) / DEC

(2)

C. Effort Estimation using Bagging Predictors

As the model is proposed to be used in early stage of development, so the following formula is used to estimate the effort.

$$\text{Effort}_{\text{nominal}} = \text{AF X (NOP)}^{B} \qquad (3)$$

The inputs are the NOP (Number of Object Points) of software development, an adjustment factor AF, and a scale factor, B. The scale (or exponential) factor, B, accounts for the relative economies or diseconomies of scale encountered for software projects of different sizes. Finally, value of factor B is adjusted using regression analysis from the data collected from all these projects. As the project size grows, it exhibits more diseconomies of scale. Further, it is found that in the maintenance based project, due to large overhead, projects exhibit more diseconomies. To improve the overall performance of the regression analysis, the bagging predictors are used. The bagging predictors' technique was first proposed by Breiman [16]. In this technique, multiple versions of same predictor over the different training sets are proposed to get an aggregated predictor. The final prediction of aggregated predictor is just the average of prediction of all the predictors. The same learning approach has to be used for each of the predictor. Main parameters of bagging approach are defined as follows:

- no-of-bootstrap-replica : As suggested by Breiman [16], we have tested our system for 25 and 50 different bootstrap replica.
- bag_percentSize –For each replica, randomly 10 projects are chosen among the 70 different students projects used for training.
- Classifier- linear regression model is used for all the replica

Different steps used for bagging predictor approach are as follows:

- Let T contains the data of all the training sets
- Initialize the variable sum to zero.
- For i = 1 to no-of-bootstrap-replica
  - Let Ti, the new training set of size bag-percentSize is created randomly from T.
  - Use the classifier to built a new regression model, Ri corresponding to Ti
  - Predict the value of each instance of T using the Ri and add to the sum.
- Compute the average of all the predictions to get the final predicted value.

Above steps are checked for two different values (25, 50) of no-of-bootstrap-replica. It is found that 25 iterations are sufficient for the bagging predictors. Results do not vary much, if we increase the number of iterations.

IV. RESULTS AND VALIDATIONS

Data is collected from the two different sets of projects. First set contains data and other details of twelve completed projects. This set is used for testing only. Similarly, second set contains data of 70 different projects undertaken by post graduate students. These student projects are used for training purpose. Different projects studied vary in the complexity of implementation of different phases of life cycle. These projects have been grouped together into the following categories using their total/build ratio parameter. These categories are created keeping CMS projects in mind.

- Type 1 - Small Projects, low total/build effort ratio
- Type 2 - Small Projects, high total/build effort ratio
- Type 3 - Large Projects, low total/build effort ratio
- Type 4 - Large Projects, high total/build effort ratio

For all these 70 projects, students have to answer two separate questionnaires, first one after studying the requirements and then the second one after completing the project. The answers of the first questionnaire are used to predict the effort using the proposed model and of the second questionnaire are used to refine the model. First questionnaire contains six different categories of questions.

Table I: Comparison of Effort Estimation using Proposed Model and Bagging Predictors

Type of Projects	No. of Projects	Mean Actual Effort	Without Using Model		Using Proposed Model		Using Bagging Predictors	
			Effort Prediction	MRE	Effort Prediction	MRE	Effort Prediction	MRE
Type 1	30	72	55	0.24	62	0.13	63	0.12
Type 2	21	82	51	0.38	69	0.16	72	0.12
Type 3	11	272	210	0.23	240	0.12	245	0.10
Type 4	18	332	219	0.34	253	0.24	290	0.13

These categories are Design requirements, Set up requirements, Implementation requirements, Usability requirements, External interface requirements, Reliability and Performance requirements, Maintainability and Adaptability requirements. Each category contains 8-10 questions about the specific requirements. Hence all the details, which are required by the proposed models, are covered in this questionnaire. Second questionnaire is used after the completion of project. In this first of all, project heads were asked about the increase in complexity as compared to their first assessment in all the different phases of software development life cycle. Secondly, they were asked to distribute the total effort in phase wise manner. The predicted effort is compared phase wise for all the projects.

The proposed model is validated using two approaches. The first approach is based on the data available from the past projects. The estimated effort of the proposed model is compared with the effort estimated in those projects. Its detailed analysis is shown in Table I. In each category, mean actual effort in person months is calculated. This is compared with the effort predicted using the proposed model and without using it. By using bagging predictors, prediction efficiency is improved as shown in Table I. Finally, the Mean Relative Error (MRE) in each category is calculated. From the Table I, it is very clear that MRE is very less when the effort is predicted using our model and further variance in prediction is improved using bagging predictors. Mean of all MRE (MMRE) using proposed model is 0.16 as compared to MMRE of 0.30 without using the model and it is further improved to 0.12 when bagging predictors are used.

The prediction at level 25 by using proposed model is more than 80%, which earlier was about 62% when the proposed model was not used.

In our second approach, the project managers were asked to estimate the effort of new project using the proposed model and another one of their choice. Then a survey is conducted using a questionnaire to check the confidence level of project manager in using the proposed model.

V. CONCLUSION

In this paper, Content Management System Effort Estimation model is proposed for the current technologies using which a piece of work can be estimated most accurately. The model is designed to help project manager to estimate effort at the very early stage of requirement analysis. The initial model is finalized using bagging predictors on linear regression model. After the requirement analysis stage, the project size metric is estimated using the object point analysis. A set of questionnaire is used to estimate the complexity of the project and the other adjustment factors, which has to be filled after completing the initial requirement analysis. This model is proposed to be used differently for the different types of projects. Different projects have been categorized based on their total/build effort ratio. After analyzing 12 completed industry projects and 70 students projects, the model shows very good accuracy for Type 1 and Type 3 projects, where total / build effort ratio is very low. But when the project contains very less coding then the effort estimation accuracy is not very good. This clearly shows that for the maintenance projects, adjustment factors needs to be modified.

REFERENCES

[1] Oreilly, tim, "What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software". Communications & Strategies, No. 1, p. 17, First Quarter 2007 http://ssrn.com/abstract=1008839

[2] Carr, L., Bechhofer, S., Goble, C., and Hall, W. "Conceptual Linking: Ontology-based Open Hypermedia". In The Tenth International World Wide Web Conference, Hong Kong, May, 2001 http://www10.org/cdrom/papers/246/index.html

[3] D. Schwabe, R. Pontes, I. Moura: "OOHDM-Web: An Environment for Implementation of Hypermedia Applications in the WWW", ACM SigWEB Newsletter, Vol. 8, #2, June 1999.

[4] Tridas Mukhopadhyay, Steven S. Vicinanza and Michael J. Prietula, "Examining the Feasibility of a Case-Based Reasoning Model for Software Effort Estimation", MIS Quarterly, Vol. 16, No. 2 June, 1992.

[5] Mendes, E.; Mosley, N.; Counsell, S., "Comparison of Web size measures for predicting Web design and authoring effort," Software, IEE Proceedings - , vol.149, no.3, pp.86-92, Jun 2002

[6] Yunsik Ahn, Jungseok Suh, Seungryceol Kim, Hyunsoo Kim, "The Software maintenance project effort estimation model based on function points", Journal of Software Maintenance & Evolution: Research and Practice, Vol 15, issue 2, April, 2003.
[7] Ruhe, M., Jeffery, R., & Wieczorek, I., "Cost estimation for Web applications", Proceedings ICSE 2003
[8] Mendes, E., Watson, I., Triggs, C., Mosley, N., & Counsell, S. , "A comparison of development effort estimation techniques for Web hypermedia applications", Proceedings Metrics 2002.
[9] Mendes, E., Counsell, S., & Mosley, N., "Measurement and effort prediction of Web applications", Proceedings of the 2nd ICSE Workshop on Web Engineering, June, 2000.
[10] Fewster, R., & Mendes, E., "Measurement, prediction and risk analysis for Web applications" Proceedings of Software Metrics Symposium, METRICS 2001.
[11] Reifer, D.J. , "Web development: Estimating quick-to-market software". IEEE Software, Nov – Dec 2000.
[12] Morisio, M., Stamelos, I., Spahos, V., Romano, D., "Measuring functionality and productivity in Web-based applications: A case study", Proceedings of the Sixth International Software Metrics Symposium, 1999.
[13] Baresi, L., Morasca, S., & Paolini, P., "Estimating the design effort for Web applications", Proceedings of Software Metrics Symposium, Metrics 2003.
[14] Baresi, L., F. Garzotto, and P. Paolini. "Extending UML for modeling web applications", In Proceedings of 34th Anual Hawaii International Conference on System Sciences (HICSS'34), Hawai, Jan. 2001.
[15] Petronio L. Barga, Adriano L.I. Oliveira, Gustavo H.T. Ribeiro, Silvio R.., "Bagging Predictors For Estimation of Software Project Effort", in Proceedings of International Joint Conference on Neural Networks, USA, August 2007.
[16] L. Breiman, "Bagging Predictors", Mach Learn, vol 24, no 2, pp 123-140, 1996.
[17] B.W. Boehm, B. Clark, E. Horrowitz, C. Westland, R. Madachy, R. Selby, "The COCOMO 2.0 Software Cost Estimation Model", American Programmer, July 1996.
[18] Vahan Harput ,Hermann Kaindl and Stefan Kramer, "Extending Function Point Analysis of Object-Oriented Requirements Specifications", Proceedings of the 11th IEEE International Software Metrics Symposium, 2005.
[19] Norman, M. Stillwell, "TeamSite Content Management Systems- Technical Overview", http://www.its.monash.edu.au/staff/web/slideshows/teamsite-technical/teamsite-tech-presentation.html
[20] English, O. Londer, T. Bleeker, S. Shell, S. Cawood, "Microsoft Contenet Management Server-2002 A Complete Guide", Published by Addison-Wesley Professional, Oct 27, 2003.
[21] Boehm, B., "Anchoring the software processes", IEEE Software, Vol 13, No 4, July 1996.

Weighted Gaussian Kernel with Multiple Widths and Network Kernel Pattern

Jing Tian[1]
Department of Electronics, School of Information Science and Engineering, Ocean University of China, Qingdao Shandong, China, 266100, kingmmx@yahoo.cn

Lifeng Zhao[2]
Department of Electronics, School of Information Science and Engineering, Ocean University of China, Qingdao Shandong, China, 266100, zhaolf@mail.ouc.edu.cn

Abstract—**Traditional Gaussian Kernel is the most widely used in Kernel Machines, such as Support Vector Machines and has been extensively studied in the field of Neural Networks, such as Radial Basis Function Network. However, because of the same weight and data distribution defined in traditional Gaussian Kernel, the weights of Neural Network based on such kernels are controlled mostly by the input data. There is no difference between the Activation Function(that is, Neuron or Neurode). We propose a new kernel called Weighted Gaussian Kernel with Multiple Widths to have more parameters to control the data. Using the new kernel in Neural Networks, we propose a new conception call Network Kernel Pattern to improve the traditional structure of Radial Basis Function Network with some new definitions.**

Index Terms: Weighted Gaussian Kernel with Mul tiple Widths, Gaussian Kernel, Neural Networks, Radial Basis Function Network, Network Kernel P attern, Pattern Factor.

I. INTRODUCTION

GAUSSIAN Kernel(GK, RBF kernel) is the most widely used kernel in Support Vector Machines(SVM)[1~3] and Radial Basis Function Network(RBFN)[4]

$$\kappa(x,y) = \exp(-\|x-y\|^2/2\sigma^2) \quad (1)$$

The feature space of GK is infinite degree. With enough numbers of GK as Activation Function(AF), the RBFN could solve all the given problems theoretically. Further more, GK has only one parameter σ to control the data and could not influence the weights of RBFN artificially, even if the weights of feature in input space should be different. Thus, we propose a new kernel based on GK called Weighted Gaussian Kernel with Multiple Widths(WGKMW)

$$\kappa(x,y) = \left[\exp(-\|x-y\|^2/2\sigma^2) + R\right]^d$$

$$\forall R \ge 0, d \ge 0, d \in N \quad (2)$$

The WGKMW have three parameters, σ, R, and d, which control the weights of feature, distribution of data in the input space and numbers of AF in RBFN.

II. NETWORK ESSENCE OF WGKMW

Consider the Binomial Theorem to unfold the expression (2)

$$\kappa(x,y) = \left[\exp(-\|x-y\|^2/2\sigma^2) + R\right]^d$$

$$\forall R \ge 0, d \ge 0, d \in N \quad (3)$$

$$= \sum_{s=0}^{d} \binom{d}{s} \left[\exp(-\|x-y\|^2/2\sigma^2)\right]^s R^{d-s}$$

$$= \sum_{s=0}^{d} \binom{d}{s} R^{d-s} \exp\left(-\frac{s\|x-y\|^2}{2\sigma^2}\right)$$

$$= R^d + \sum_{s=1}^{d} \binom{d}{s} R^{d-s} \exp\left(-\frac{\|x-y\|^2}{\frac{2\sigma^2}{s}}\right)$$

$$1 \le s \le d, s \in N, d \in N \quad (4)$$

Define:

$$\psi(x) = \sum_{s=1}^{d} \binom{d}{s} R^{d-s} \exp\left(-\frac{\|x-c_s\|^2}{\frac{2\sigma^2}{s}}\right)$$

$$1 \le s \le d, s \in N, d \in N \quad (5)$$

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_6, © Springer Science+Business Media B.V. 2010

where c_s =y, are the centers of input data set and could have the same values. Further more, Define:

$$a_s = \binom{d}{s} R^{d-s} \tag{6}$$

$$\rho(\| x - c_s \|) = \exp\left(-\frac{\| x - c_s \|^2}{\frac{2\sigma^2}{s}}\right) \tag{7}$$

Thus, the expression (5) could be expressed in a more general way[5]

$$\psi(x) = \sum_{s=1}^{d} a_s \rho(\| x - c_s \|)$$

$$1 \le s \le d, s \in N, d \in N \tag{8}$$

where d is the number of neurons in the hidden layer of Neural Network(NN), c_s is the center vector of neuron, a_s is the linear weight of linear output in NN, and $\rho(\| x - c_s \|)$ is the nonlinear AF of NN, all of which construct the framework of RBFN. The essence of WGKMW is a certain kind of Gaussian RBFN(GRBFN)[6] but has great difference comparing to traditional GRBFN.

III. NETWORK KERNEL PATTERN

Note that compared to the traditional GRBFN, whose weights are completely determined by the Objective Function(OF) and certain iterating algorithm such as Gradient Descent Algorithm(GDA), in WGKMW, the weights a_s is also influenced by parameter R and d.
Define:

$$\alpha(s) = \binom{d}{s}$$

$$1 \le s \le d, s \in N, d \in N \tag{9}$$

This disrete function is symmetric in middle value u, where $u = d/2$ when d is even and $u = (d \pm 1)/2$ when d is odd, with maximum when s=u and minimum 1. In Fig. 1, we show the distribution of $\alpha(s)$ when d=7.

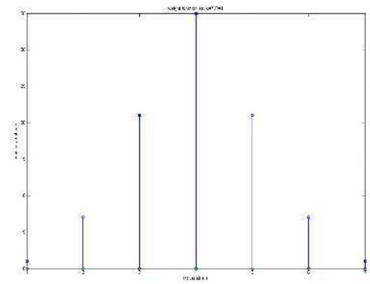


Fig. 1. $\alpha(s)$ distribution when d=7

Define:

$$\beta(s) = R^{d-s} \tag{10}$$

This discrete function is controlled by R, which is decreasing and concave, gets the minimum 1 at s=d, when R>1, which is increasing and concave, gets the maximum 1 at s=d, when 0<R<1. $\beta(s) \equiv 1$, when R=1. In Figs. 2 and 3, we show distribution of $\beta(s)$ when R=1.5 and R=0.5, both d=7.

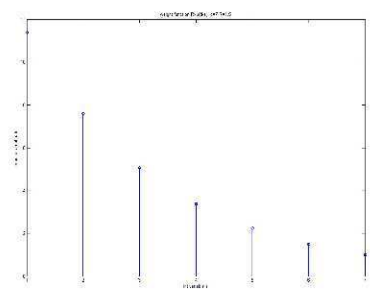


Fig. 2. $\beta(s)$ distribution when R=1.5, d=7.

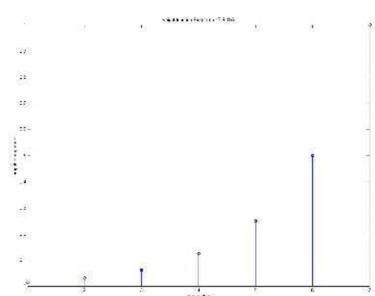


Fig. 3 $\beta(s)$ distribution when R=0.5, d=7.

Define:

$$a_s = \alpha(s)\beta(s) \tag{11}$$

This is the weight function of WGKMW, which is increasing and local convex(not strict globle convex, that is meaning on certain successive values depending on the parameters of s and R.), when s<u and R>1, and is decreasing and local concave when s>u and R>1. The situation of 0<R<1 is that the function is increasing and local concave when s<u and decreasing and local convex when s>u. In Figs. 4 and 5, we show the distribution of a_s when R=1.5 and R=0.5, both d=7.

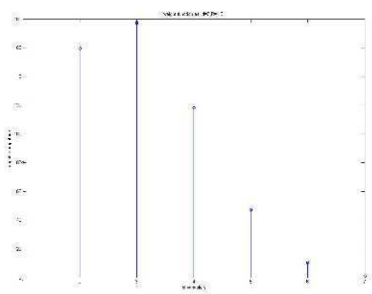


Fig. 4. a_s distribution when R=1.5, d=7

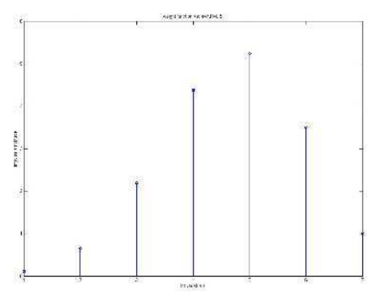


Fig. 5. a_s distribution when R=0.5, d=7.

When R=1, the distribution of a_s is the same as expression (9) as Fig. 1.
It is clearly that the parameter R controls the weigths distribution with the given number of neurons. When R>1, the first Uth neurons contributes more(that is, with bigger weight power). When R=1, the middle part of neurons take an important role. When 0<R<1, the last Uth neurons makes the point.

Note that $\rho(\| x - c_s \|) = \rho(s)(\| x - c_s \|)$, which means that all the AF(here is GK)in the RBFN have different varience, which is changed from σ^2, $\frac{\sigma^2}{2}$,…, $\frac{\sigma^2}{s}$,…to $\frac{\sigma^2}{d}$.

Comparing to traditional RBFN based on GK, WGKMW has Relative Weights Power(RWP) which is controlled by parameter R, and a series GK as AF but with linear changeable varience. In Fig. 7, we show the WGKMW structure comparing with traditional GRBFN in Fig. 6.

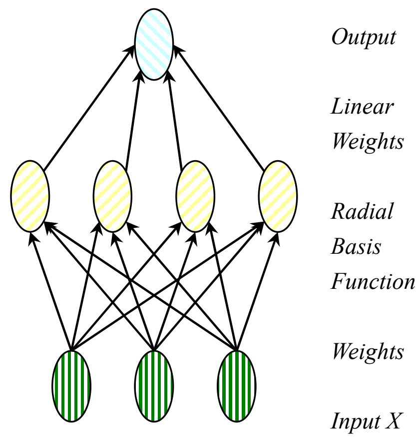


Fig. 6. traditional structure of RBFN

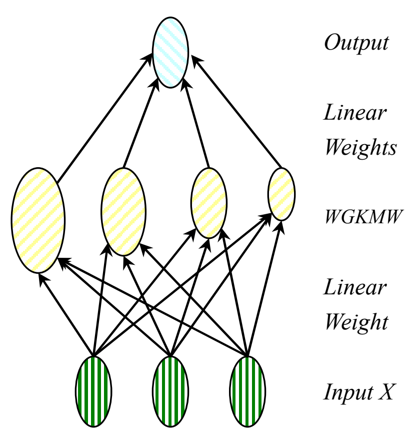


Fig. 7. new structure of WGKMW

Based on WGKMW, we propose a new conception—Network Kernel Pattern (NKP), if the function is satisfied with expression(12):

$$\mathrm{N}(x) = [\rho \| x - c_s \| + R]^d$$

$$\forall R \ge 0, d \ge 0, d \in N \tag{12}$$

where $\rho(\| x - c_s \|)$ is any radial basis function defined in the RBFN. And R is the Pattern Factor(PF), which controls the RWP of the RBFN.

IV. RBFN BASED ON NKP

Consider that a NKP is not only an independent RBFN, but also is an AF in the NN. When certain NKP is embeded into the RBFN, a new structure of RBFN with two hidden layers is created.

$$\psi_{\mathrm{N}}(x) = \sum_{i=1}^{n} \beta_i P(\| x - c_i \|) \tag{13}$$

$$= \sum_{i=1}^{n} \beta_i \mathrm{N}(x) = \sum_{i=1}^{n} \beta_i [\rho(\| x - c_i \|) + R]^d$$

$$= \sum_{i=1}^{n} \beta_i [R^d + \sum_{s=1}^{d} \binom{d}{s} R^{d-s} \rho^s (\| x - c_i \|)]$$

$$= \sum_{i=1}^{n} \beta_i R^d + \sum_{i=1}^{n} \beta_i \sum_{s=1}^{d} \binom{d}{s} R^{d-s} \rho^s (\| x - c_i \|) \tag{14}$$

In Fig. 8, we show the characteistics of such new structure of RBFN when a NKP is used.

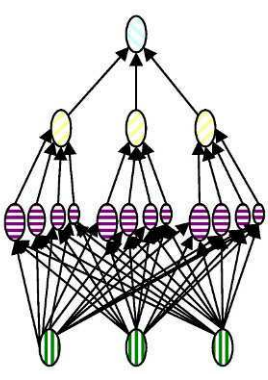


Fig. 8. structure of RBFN based on NKP

Note that the communication(that is, information transition) between the 1st and the 2nd hidden layers is different with the traditional RBFN which has two hidden layers. The neurons in the 1st layer with the same centers could only influence only one neuron in the 2nd layer. That is neurons with different centers in the 1st layers could not add weights to the same neuron in the 2nd layer.

V. CONCLUSION

In this work, we have proposed a new kernel based on GK called WGKMW. With more parameter to change than GK, WGKMW could have more flexibility. More important, we have developed the essence of WGKMW, that is a kind of RBFN, from which we have proposed a new conception called NKP based on WGKMW. Finally, we have shown the general form of RBFN based on NKP, which displays a new structure of hidden layers.

REFERENCES

[1] V. Vapnik, "Universal Learning Technology: support vector machines"[J], NEC Journ of Adv Tech, pp. 137-144, Feb, 2005.
[2] V. Vapnik, "The Nature of Statistical Learning Theory"[M], New York, Springer-Verlags, 1995.
[3] V. Vapnik, "Statistical Learning Theory"[M], New York, Wiley, 1998.
[4] M. Powell, "Radial basis functions for multivariate interpolation", in Algorithm for Approximation[J], Oxford, Clarendon Press, pp. 143-167, 1987.
[5] S. Haykin, Neural Networks: A Comprehensive Foundation. Upper Saddle River, N.J., Prentice Hall, 1999, 2nd Edition.
[6] M. Figueiredo, "On Gaussian Radial Basis Fucntion Approximation: Interpolation, Extenstions, and Learing Strategies", 1049-001 Lisboa, Portugal, 2002.

New Advanced Graduate Course Designed to Provide Students Exposure to Today's Developing and Challenging World

Zheng (Jeremy) Li, Ph.D.
Associate Professor, School of Engineering
University of Bridgeport, Bridgeport, CT 06604, USA

Abstract

The rapidly evolving engineering technology and globalization process have changed educational roles and expectations from teachers. All these lead to some curriculum and teaching reforms that focus on introducing potential programs which provide students with challenging technologies. A recent graduate course "Production Techniques and Technologies" has been developed at University of Bridgeport to establish entrepreneurship opportunities for fast-pace and developing world. Some advanced technologies, including Failure Modes and Effects Analysis (FMEA), Standard Operation Procedure (SOP), and Lean Manufacturing, have been introduced in helping students to gain high level of engineering experiences. It also helps students to examine political, social and economic issues that associated to engineering project delivery processes that can lead to save product process expenses, increase manufacturing efficiency, improve product quality and shorten production period. Although this course was taught at mechanical/manufacturing engineering majors, it also can be applied to most areas of US industrial and engineering practices.

Introduction

Many students enrolling into the engineering major will have the difficulty to choose an appropriate engineering discipline. One of the good ways is to cover the disciplines through multidisciplinary engineering studies and provide students with the real industrial case study and feasible/potential solutions. The explanation of how engineers in different disciplines can contribute to the solution of the engineering issue could become the feasible ways of obtaining a general idea of what is the duty of engineers in each engineering discipline. Related engineering examples from industry are good resources in the engineering classroom, and it can be applied to help students to understand the real issue and to show how the technical concept can be used in real cases.

The Failure Model and Effects Analysis (FMEA) has been applied to many different industries through the world for years and implementation of this technology will save product process expenses, increase the manufacturing efficiency, improve the product quality and shorten the production period which will enhance the company's competition in the market. To understand the potential failures and effects on companies' daily production and manufacturing processes, several modeling and simulation tools will be used to help understanding the major factors that affect the manufacturing processes, process efficiency, product quality and production quantity. FMEA is intended to document:

(1). A failure

(2). Its mode

(3). Its effect

(4). By analysis with focus on lean Manufacturing

The basic idea of Failure Model and Effects Analysis (FMEA) is to spot risks and to initiate dedicated efforts to control or minimize risks in manufacturing processes. Knowing your risks can make your production and manufacturing process plan more realistic. FMEA seems to work best, when an engineering team documents its known knowledge about known cause- and effect-relationships. In this analysis mode, the timely sequence of failure events should be worked out first, before entering results into a FMEA sheet. The Ishikawa or flow diagrams can be introduced to organize the items into a logical sequence of events and start with key functions your manufacturing process/production has to perform and then identify ways to deteriorate those key functions (modes and causes). It is not easy to anticipate the unknown risks and unknown failure mechanisms in manufacturing

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_7, © Springer Science+Business Media B.V. 2010

process/production and AFD (Anticipatory Failure Determination) will be recommended when you need to analyze subtle failure mechanisms or have to predict future failures from the manufacturing process/production. It is important and critical to stay well disciplined in cause- and effect-thinking to create strong FMEA-results.

Lean Manufacturing can improve the industrial material handling, inventory, quality, scheduling, personnel and customer satisfaction. With the Standard Operation Procedures (SOP) in manufacturing, and ISO 9001/ISO14000 standard, the plant performance in manufacturing flow, organizational functions, process control, metrics and logistics can be improved.

The feedback from our many graduated students who are currently working in different industries indicated that all these knowledge learnt from our class are important for them to be succeeded in their career to face today's challenging world. These techniques will also determine if the businesses are successful or not. This shows that, if we educate the important technologies to our students early in the school, they can get ready to face the challenge when they leave the school.

Case Study and Projects

Several examples from real engineering practices are discussed in the class to help students to understand the concepts of these hot technologies. The thorough discussion on these real case studies will indicate how the feasible solutions can improve the quality of engineering education in today's engineering school. Two of the class case studies of applying the FMEA are as follows:

Case 1:

Function	CAM shaft rotation
Failure Mode	CAM follower never rotates
Effects	Machine is not working
S(severity rating)	9
Cause(s)	CAM shaft coupling loose
O(occurrence rating)	3
Current Controls	Inspection of CAM coupling assembly
D(detection rating)	4
CRIT(critical characteristic)	N
RPN(risk priority number)	108
Recommended actions	Performing tolerance and structure analysis
Responsibility and target completion date	Eric Smith 15-Oct-2006
Action taken	

Case 2:

Function	Filling tub
Failure Mode	Control pressure sensor never trips
Effects	Water spills on user floor
S(severity rating)	7
Cause(s)	Pressure control sensor malfunctioned Pressure sensor is not proper connected
O(occurrence rating)	4
Current Controls	Fill timeout based on time to fill to low pressure sensor
D(detection rating)	3
CRIT(critical characteristic)	N
RPN(risk priority number)	84
Recommended actions	Perform cost analysis of adding additional sensor halfway between low and high pressure sensors
Responsibility and target completion date	Janis Clinton 10-Nov-2005
Action taken	

Here all failure modes should be determined by students based on the functional requirements and their effects. A failure effect is considered as the result of a failure mode on the function of the system as perceived by the customer. So it is convenient to write these effects down in terms of what the customer might see or experience. The examples of failure effects include: degraded performance, noise or even injury to a customer. Several important factors in the above cases can be detailed as follows:

(1) Severity: Each effect will be assigned a severity number (S) from 1 (no danger) to 10 (important) that can help all engineers to identify the defect levels. If the severity of an effect is 9 or up, necessary actions are considered to modify the design by eliminating the failure mode to protect customer from the effect. The severity rating of 9 or 10 is generally considered for those effects which will cause injury to a customer or otherwise result in a lawsuit.

(2) Occurrence: It is important to investigate the cause of a failure and determine how many times it can occur. This can be done by comparing the similar products or processes and the failures that have been previously documented. A failure cause is considered as a design weakness and all the potential causes for a failure mode must be verified and documented. A

failure mode is given a probability number (O) with number 1-10. Actions must be taken if the occurrence is high (i.e. > 4 for non safety failure modes, and > 1 when the severity number is 9 or up). This step is also considered the detailed development section of the FMEA process.

(3) Detection: While the proper actions are determined, it is necessary to verify their efficiency and the design verification should be considered. The appropriate inspection techniques need to be chosen. The engineers should first review the current controls of the system that either prevent failure modes from occurring or detect the failure before it reaches the users. Engineers should determine the testing, analysis, monitoring and other techniques that will be or have been used on similar systems to detect failures, and learn how likely the failures can be identified or detected from these controls. The detection number (D) represents the ability of planned tests and inspections to remove defects or detecting failure modes.

(4) Risk Priority Numbers: Based on above 3 basic steps, the Risk Priority Numbers (RPN) can be calculated. RPN are not important in the choice of an action against failure modes but they are more threshold values in the evaluation of these actions. After determining the numbers of severity, occurrence and detection, the RPN can be easily calculated by: RPN = S x O x D

This procedure must be observed for the entire project or production process. Once this is done it will be easy to verify the areas of greatest concern. The failure modes with the highest RPN must be given the highest priority for corrective action. This indicates that it is not always to first treat the failure modes with the highest severity numbers. There could be less severe failures, but with more occurrences and lower detection.

After these values are determined, some recommended actions with targets, responsibility and dates of implementation are documented. These actions can include special inspection and quality testing. Once the actions have been implemented in the design/process, the new RPN must be re-checked again to confirm the positive improvements. If a design or a process changes, an FMEA should be simultaneously updated. In order to improve the current engineering process based on FMEA, the severity of the failure should be minimized, the occurrence of failure mode must be reduced, and the detection methods ought to be improved.

Through these case discussions, the students will understand that, in industrial and engineering processes, FMEA can provide an analytical approach while dealing with potential failure modes and their associated causes. When considering possible failures in a design, such as safety, cost, performance, quality and reliability, the engineers can gain many information on how to change the development/manufacturing processes to prevent these failures from happening as best as possible. FMEA introduces an efficient and easy tool to verify which risk has the greatest concern, and which correct action is needed to prevent a problem before it arises. The development of these specifications will ensure the product to meet the required functions.

All class formal projects, which are usually selected from some U.S. famous and successful companies, are introduced and assigned to students to help them in learning and establishing of industrial & engineering standards in today's industry. One of the class projects is as follows:

The workshop manager walked into the workshop and found some defect product in the shop floor. He called the foreman over and told him to find out what cause the product problem. Next day while the workshop manager was in the same area of the workshop, he found other defect products again and he subsequently criticizes the foreman over the defect products for not following his directions from the day before. Please use the FMEA technique and Lean Manufacturing concept to improve the shop production situation.

Conclusion

Today's fast-path developing and challenging world requires college students/future engineers to gain more advanced and professional engineering knowledge, in order to be successful in their future career. The teaching reform efforts in this paper focus on directly improving faculty teaching capability and student learning curve to face today's challenging world. A new course, currently taught at the mechanical/manufacturing engineering majors, was designed to introduce professional issues associated with today's US engineering practice. This new course places more emphasis on how to guide engineering students to gain the necessary industrial and engineering knowledge, that are usually missing from the current engineering course curriculum, including basic concept of FMEA, Lean Manufacturing, Six Sigma, ISO 9001 and ISO 14000. In addition to the numerous real industrial cases' study and discussion, a number of real projects have been developed in this course to help and allow students to actively apply the skills learnt from the lecture. The students will be benefited from these projects by interacting in a group environment with other multidisciplinary students. Finally, the students will learn about the industrial environment in the developing world and come up with feasible and sustainable solutions to today's industrial and engineering issues. All these hot and useful industrial knowledge, combined with traditional engineering technology taught in current MS engineering degree level, can help students in their soon job hunting process and future engineering career. The feedback from our many graduated students confirms the promising curriculum reform in this new graduate engineering course.

References

[1]. P. Kim, T. Miranda and C. Olaciregui, "Pocket School: exploring mobile technology as a sustainable literacy education option for underserved indigenous children in Latin America", International Journal of Educational Development 28 (4), pp. 435–445, 2008.

[2]. G.M. Grossman, P.E. Onkol and M. Sands, "Curriculum reform in Turkish teacher education: attitudes of teacher educators towards change in an EU candidate nation", International Journal of Educational Development 27, pp. 139–150, 2007.

[3]. G. Grossman, P. Onkel and M. Sands, "Curriculum reform in Turkish teacher education: attitudes of teacher educators towards change in an EU candidate nation", International Journal of Educational Development 27 (2), pp. 138–150, 2007.

[4]. P.K. Kubow and P.R. Fossum, "Comparative Education: Exploring Issues in International Context", Pearson Education/Merrill Prentice Hall, Upper Saddle River, NJ 2007.

[5]. V. Masemann, M. Bray and M. Manzon, "*Common Interests, Uncommon Goals: Histories of the World Council of Comparative Education Societies and its Members*, Comparative Education Research Centre", The University of Hong Kong, Hong Kong, 2007.

[6]. M. Bray, B. Adamson and M. Mason, "*Comparative Education Research: Approaches and Methods. CERC Studies in Comparative Education",* vol. 19, Comparative Education Research Centre, The University of Hong Kong, Hong Kong, 2007.

[7]. D. Phillips and M. Schweisfurth, "Comparative and International Education: An Introduction to Theory, Method and Practice", Continuum, London 2006.

[8]. A. Aypay, "Re-structuring higher education: the influence of socio-economic and political environments to institutional adaptation of Universities", *Eğitim Yönetimi* 34, pp. 194–213, 2003.

[9]. Okcabol, R., "Teacher qualifications and the new council of higher education model", II. Ulusal Öğretmen Yetiştirme Sempozyumu, 10–12 Mayıs, Çanakkale, Turkey, pp. 419–425, 2000.

Dynamic Fire 3D Modeling Using a Real-time Stereovision System

L. Rossi[1], M. Akhloufi[2]

[1] UMR CNRS SPE 6134, University of Corsica, 20250 Corte, France
[2] CRVI, 205, rue Mgr-Bourget, Lévis (Québec), G6V 6Z9, Canada

Abstract – **This work presents a new framework for three-dimensional modeling of dynamic fires present in unstructured scenes. The proposed approach addresses the problem of segmenting fire regions using information from YUV and RGB color spaces. Clustering is also used to extract salient points from a pair of stereo images. These points are then used to reconstruct 3D positions in the scene. A matching strategy is proposed to deal with mismatches due to occlusions and missing data. The obtained data are fitted in a 3D ellipsoid in order to model the enclosing fire volume. This form is then used to compute dynamic fire characteristics like its position, dimension, orientation, heading direction, etc. These results are of great importance for fire behavior monitoring and prediction.**

I. Introduction

Research in the field of fire needs tools that permit to track fire characteristics over time. With the obtained results, we can compare the theoretical models to the situation in the ground. Also, in many real situations information about fire, like its position, orientation, height and volume, are crucial for firefighting. Examples of these situations are: compartments fire [1], propagation fire [2], oil platform fire [3].

For more than two decades, visual and infrared cameras have been used as complementary metrological instruments in fire and flame experiments [4-8]. Vision systems are now capable to reconstruct a 3D turbulent flame and its front structure when the flame is the only density field in the images [9, 10]. Image processing methods are also applied to follow forest fire properties like the rate of spread of flame fronts, fire base contour, flame orientation and maximum flame height [11-14]. These techniques extract information from a scene using multiple viewpoints (in general, frontal and lateral views) and synthesize the data in subsequent steps. This type of processing cannot be used in all experimental situations.

In computer vision, many works has been done for the reconstruction of 3D object shape. However, the majority of these works deal with rigid objects [15-17]. Little work has been done for modeling 3D non rigid objects and many hypotheses are made in order to achieve an acceptable 3D reconstruction in this case [18-20]. The lack of a framework for 3D reconstruction of complex fire fronts is due to the dynamic and random nature of the fire which makes it difficult for vision systems to efficiently locate and match salient points.

In this paper, we present a new framework for a three dimensional modeling of dynamic fires. This framework is robust to outdoor unstructured conditions. It permits the extraction of important 3D data and models the fire by means of a 3D ellipsoid. This later shape permits the extraction of important fire characteristics, like its dimensions, angles and volume.

II. Proposed approach

The proposed framework uses stereovision for three-dimensional modeling of parts extracted from fire regions. In order to build 3D data from stereo, we need to extract corresponding points in the left and right images of the stereo pair. The steps involved in the proposed approach are:

1. Segmentation of fire images in order to extract fire regions.
2. Features detection algorithm for extracting salient points from the segmented regions.
3. Best features selection using correlation based matching strategy. This step permits the refinement and selection of salient points and the construction of a set of corresponding points.
4. Computation of three-dimensional fire points using stereo correspondence.
5. 3D Ellipsoid fit for volume reconstruction and fire characteristics computation.

III. Vision Framework For Three Dimensional Modeling of Fire

A. Stereo Vision

Since our objective is to model three-dimensional shapes of fires in operational scenarios, we need a system than can be deployed quickly and efficiently in unstructured scenes. A stereo camera pair is the best choice in this case. A stereo vision system relies on epipolar geometry constraints to derive the 3D structure form 2D correspondences [21, 22-24]. Stereo triangulation permits to extract the 3D coordinates of a point form its 2D projections in two or more images. 3D position is depicted by stereo disparity, defined by the difference in the position of the projected point in two images.

In order to compute the 3D position from 2D correspondences, the system must be calibrated. Calibration allows the calculation of cameras intrinsic and extrinsic parameters. Since calibration can be a difficult task in an outdoor scenario, we chose to use a pre-calibrated camera. With this type of camera, there is no need for calibration. A Point Grey XB3 Trinocular stereo system [25] was used in our experiments.

In stereo processing, image rectification is an important step. Image rectification permits to align epipolar lines in the

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_8, © Springer Science+Business Media B.V. 2010

two stereo images. Corresponding point in the right image are then located within the line passing through the same "Y" coordinate as their left counterpart [21, 22-23].

In order to derive the 3D data from the obtained images, we need to find corresponding points in the left and right images. Since fire is of dynamic and random nature, we have developed a new approach to extract salient corresponding points present in the two stereo images. The following sections give more details about the new technique.

B. Two levels segmentation

The fire image is characterized by dominant colors in yellow, orange and red intervals. Also, color variations inside the fire flame give rise to homogeneous color regions (Fig. 1). These characteristics will be used in the proposed two-level segmentation technique. The proposed approach is robust and can handle large variations present in unstructured scenes

1) First level segmentation: a combination of YUV and RGB information

Work we have conducted using different color spaces for a better segmentation of fire regions in complex unstructured scenes showed that "V" channel of the YUV color system [26-27] is interesting for finding fire regions. However, in outdoor scenes other areas not corresponding to fire can appear close to fire regions (Fig. 2). K-means clustering technique is applied to "V" channel in order to extract the most interesting areas corresponding to fire.

K-means clustering is used to find *k* partitions based on a set of attributes. It permits to find iteratively the centers of natural clusters in the data. It assumes that the object attributes form a vector space. The objective of the algorithm is to minimize total intra-cluster variance:

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} \left\| x_j - m_i \right\|^2 \tag{1}$$

Where:
k : is the number of clusters.
S_i : are the clusters $i \in \{1, \ldots, k\}$.
m_i : is the centroid of all the points $x_j \in S_i$.

Fig. 1. Original image.

Fig. 2. "V" channel of YUV fire image

We used K-means clustering technique with k = 2 (1 cluster for the fire regions and 1 cluster for the background). Fig. 2 shows the results of this processing.

Assuming that fire region corresponds to the bigger extracted zone in the clustered "V" channel, we used a learning strategy in the RGB color space to compute a reference model for color classification of fire pixels.

We use a 3D Gaussian model to represent the pixels present within the fire zone. The 3D Gaussian model is defined by the following mean and standard deviation:

$$\begin{aligned} \overline{m} &= (m_R, m_G, m_B) \\ \sigma &= \max(\sigma_R, \sigma_G, \sigma_B) \end{aligned} \tag{2}$$

Where:
m_i : is the mean for channel *i*.
σ_i : is the standard deviation for channel *i*.
$i \in \{R, G, B\}$

The pixels present in the white clustered "V" channel are then verified in the RGB image in order to see if their colors are close to the reference fire color.

A pixel is classified based on the model learned from a reference area. A pixel is represented by a 3D vector defined by its color components: $p = (p_R, p_G, p_B)$. It is classified using the following formulation:

$$\begin{cases} \left\| p - \overline{m} \right\| \le k \times \sigma & z \in Fire \\ Otherwise & z \notin Fire \end{cases} \tag{3}$$

Where:
$\left\| p - \mathrm{m} \right\| = [(p_R - m_R)^2 + (p_G - m_G)^2 + (p_B - m_B)^2]^{1/2}$
k : is a constant.

The result of this first segmentation is a fire region as shown in Fig. 3.

Fig. 2. Extracted fire region.

2) Second level clustering and segmentation

After the first level segmentation we obtain a fire region with different homogeneous color zones. A second level segmentation is then performed in the resulting image. K-means clustering technique is used to extract the interior fire regions. Fig. 4 shows the result of clustering the fire image in 4 clusters (3 clusters for the fire regions and 1 cluster for the background).

C. Detection of feature points

1) Contour Extraction.

The obtained image in the first level segmentation is used to extract the global contour of the fire region. The extracted region is binarized. The obtained binary image highlights abrupt discontinuities present in the image. A postprocessing step based on mathematical morphology is then conducted in order to eliminate spurious pixels, such as residual burning embers, and to correct imperfect segmentation results like holes that appear in the fire area due to the presence of smoke. The final contour is then obtained using Canny edge detection algorithm. The result is a list of points representing the bordering pixels along the global fire region (Fig. 5 (a)).

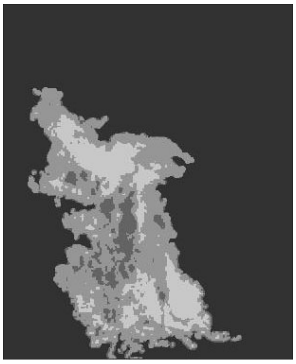


Fig. 4. Clustering of interior fire regions.

The two-level segmentation permits the labeling of interior homogeneous color regions. Theses regions are then processed separately in order to extract their contours. A labeled region is extracted in another image and binarized. Canny edge detection algorithm is then applied in order to extract the labeled region contour (Fig. 5 (b)).

2) Features detection.

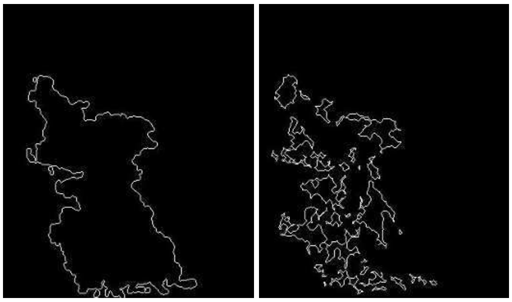
Features like edges and textures are not easily found in fire images. In our approach we use peaks and valleys of the fire contour as feature points. A peak detector is used in order to find local positive and negative inflections along the extracted contour denoted $f(x)$:

$$\begin{cases} f(x_i) < f(x_j) \Rightarrow Minima & j \in \{i-1, i+1\} \\ f(x_i) > f(x_j) \Rightarrow Maxima & j \in \{i-1, i+1\} \end{cases} \tag{4}$$

Fig. 6 shows the result obtained with the peak detector: 818 points where extracted from the left image and 830 points from the right image.

3) Matching and refinement of features selection.

The previous step permits the extraction of all feature points satisfying our extrema detection criteria. Not all of these points can be matched due to occlusion and local color variations.



(a) (b)

Fig. 5. Contour extraction: (a) Global contour of the fire region; (b) Interior fire regions contours.

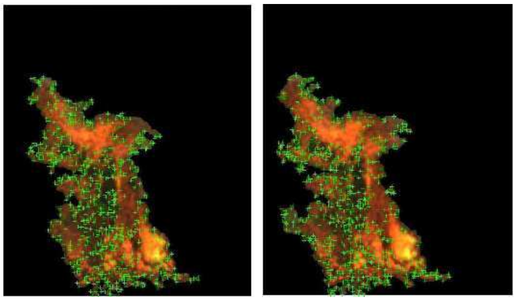


Fig. 6. Features detection.

The matching procedure selects the best features and finds corresponding points in the left and right images. The algorithm uses the following constraints during the matching process: epipolar, order and uniqueness constraints. Also, we have added a disparity constraint that restricts the search in a small interval along the epipolar search line.

For each detected feature in the rectified left image, we start a 1D search for its corresponding feature point in a small area along the horizontal line in a rectified right image (epipolar and disparity constraints). The search algorithm uses a normalized cross-correlation pattern matching algorithm in a 33x33 window around each potential corresponding point. The similarity criterion is the correlation score between windows in the two images [20]. When two possible correspondences to the left point are present in the right image and have close matching scores, priority is given to the leftmost unmatched point (order and uniqueness constraints).

Fig. 7 shows the remaining points after the matching procedure: 281 corresponding points are detected.

D. 3D positions

Once the corresponding features are extracted, a triangulation technique is used to compute their 3D coordinates. A line l is passed through the point in the left image and the left optical center C_l and a line r is passed through the corresponding point in the right image and the right optical center C_r. A mid-point method finds the point which lies exactly at the middle of the shortest line segment which joins the two projection lines. This point represents the 3D coordinate of the corresponding pixels.

E. Ellipsoid modeling

From the computed 3D points, we use the Khachiyan's algorithm [28] to obtain the minimum-volume enclosing ellipsoid. The radii and orientation of the ellipsoid were obtained by decomposing the ellipsoid matrix with the Singular Value Decomposition (SVD). This information was used to approximate the position, dimensions, orientation and volume of the fire shape model. Fig. 9 shows the ellipsoid obtained corresponding to the fire presented in Fig. 1.

Fig. 7. Refinement of selected features and matching.

Fig. 8 shows the sparse 3D representation of the obtained points.

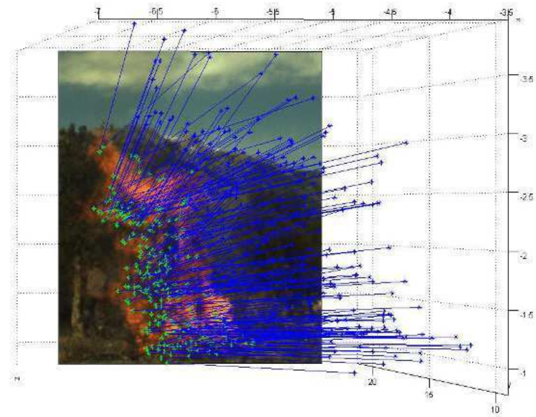


Fig. 8. 3D position of corresponding points.

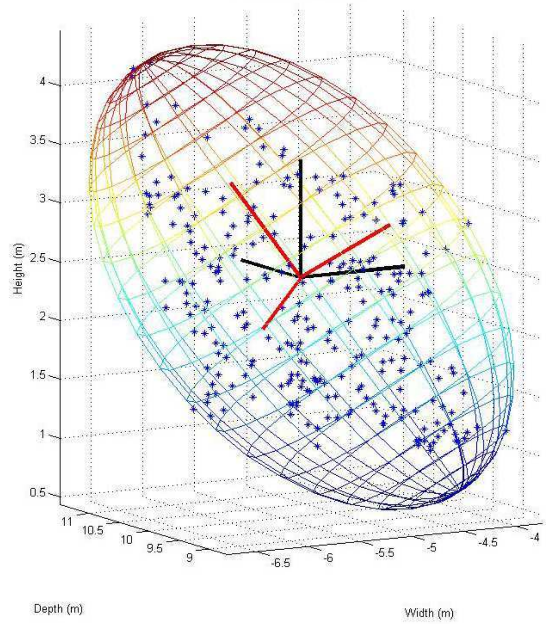


Fig. 9. Ellipsoid modeling of fire

IV. Experimental Results

Experiments were conducted in an operational scenario: outdoor unstructured environment. Tests were conducted outdoors during the day in Corte region (Corsica Island, France). Since the fire has a dynamic and random nature, the acquisition of the two images must be synchronized. We chose to use a pre-calibrated stereo camera form Point Grey (XB3 Trinocular stereo system). This camera gives perfectly registered images (synchronization time is around 125µs). The images were captured at full resolution 1280x960. Brightness and white balance were adjusted before acquisition and integration time was set to 0.1ms to avoid successive image averaging. The stereo system was positioned at approximately 12m from the fire front. The height of the fire was approximately 2.5m. The obtained results of fire 3D model are in concordance with these data.

Captured images were stocked in RAW format and later processed for segmentation and 3D reconstruction. Image acquisition code was developed in C++ and optimized for real time processing. Image processing was done offline using Matlab. Fig. 10 shows the experimental setup used for capturing fire fronts in outdoor unstructured scenes. The arrow points to the XB3 stereo system. The figure shows other cameras used in our experiments (visible and infrared: NIR, LWIR). These cameras are part of ongoing research in multi-spectral processing of fire images.

V. Conclusion

In this work we present a new framework for robust 3D modeling of dynamic fires. The proposed approach deals with various problems arising from the dynamic nature of fires like occlusions, changing colors, etc. A two-level segmentation technique was introduced. The first level starts with a global segmentation of fire regions from the unstructured scene, and the second level segments the fire image in multiple interior regions of corresponding colors. The two levels use clustering techniques. A new matching strategy is also presented. It handles efficiently mismatches due to occlusions and missing data. In order to obtain fire shape, an ellipsoid modeling based on the obtained 3d data is proposed. From this model, information like position, orientation, dimensions and volume of fire are computed. These data can be used efficiently in operational scenarios for tracking and predicting fire fronts propagation.

Future work includes the fusion of different strategies for enhanced detection of feature points and the use of multi-spectral images in order to add more information to the processing steps. Also, of importance is the final rendering of the 3D data. Finally, work will be conducted to optimize the image processing for real time.

Fig. 10. Experimental setup.

This step will permit the deployment of the proposed setup in operational scenarios during fire fighting in conjunction with the fire spread models.

Acknowledgment

The present work was supported in part by the Corsican region under Grant ARPR 2007-2008.

References

[1] C. Abecassis-Empis et al., “Characterisation of Dalmarnock fire Test One”, in *Experimental Thermal and Fluid Science*, 2008.

[2] J. H. Balbi, J. L. Rossi, T. Marcelli, P. A. Santoni, “A 3D physical real-time model of surface fires across fuel beds”, in *Combustion Science and Technology*, vol. 179, 2007, pp. 2511-2537.

[3] J. Huseynov, S. Baliga, A. Widmer., Z. Boger, “2008 Special Issue: An adaptive method for industrial hydrocarbon flame detection" in *Neural Networks* 21(2-3): 398-405, 2008.

[4] C. M. Britton, B. L. Karr, F. A. Sneva, “Correlation of weather and fuel variables to mesquite damage by fire”, in *Journ. of Range Manag.*, vol. 30, 1977, pp. 395-397

[5] H. B. Clements, “Measuring fire behavior with photography”, *Photogram. Engineer. And Remote Sensing*, vol. 49, n°10, 1983, pp. 1563-1575.

[6] G. Lu, Y. Yan, Y Huang and A. Reed, “An Intelligent Monitoring and Control System of Combustion Flames”, *Meas. Control*, vol. 32, n°7, 1999, pp. 164--168.

[7] D. X. Viegas, M.G.Cruz, L.M. Silva, A.J. Ollero, et al., “Forest Fire Research and Wildland Fire Safety”, *Proc. of IV Intern. Conf. on Forest Research and Wildland and Fire Safty Summit*, 2002.

[8] H. C. Bheemul, G. Lu and Y. Yan, “Digital Imaging Based Three-Dimensional Characterization of Flame Front Structures in a Turbulent Flame”, *IEEE Trans. on Instr. And Measur.*, vol. 54, 2005, pp. 1073-1078.

[9] G. Lu, G. Gilabert and Y. Yan, “Three Dimensional Visualisation and Reconstruction of the Luminosity Distribution of a Flame using Digital Imaging Techniques”, *Journ. of Physics: Conference Series*, vol. 15, 2005, pp. 194-200

[10] S. W. Hasinoff and K.N. Kutulakos, “Photo-Consistent Reconstruction of Semitransparent Scenes by Density-Sheet Decomposition”, IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 29, n°5, 2007, pp. 870-885.

[11] J. R. Martinez-de Dios, J.C . André, J.C.Gonçalves, B.Ch. Arrue, A. Ollero and D.X. Viegas, “Laboratory Fire Spread Analysis Using Visual and Infrared Cameras”, Inter. Journ.of Wildland Fire, vol. 15, 2006, pp.175-186.

[12] E. Pastor, A. Àgueda, J. Andrade-Cetto, M. Muñoz, Y. Pérez, E. Planas, “Computing the rate of spread of linear flame fronts by thermal image processing”, Fire Safety Journal, vol. 41, n°8, 2006, pp. 569-579.

[13] L. Merino, F. Caballero and J.R. Martínez-de Dios, J. Ferruz, A. Ollero. “A cooperative perception system for multiple UAVs: Application to automatic detection of forest fire”, *Journal of Field Robotics*. vol. 23, n°3, John Wiley, 2006, pp. 165 - 184.

[14] J. R. Martinez-de Dios, B.C. Arrue, A. Ollero, L. Merino, F. Gomez-Rodriguez, “Computer vision techniques for forest fire perception”, *Image and vision computing,* vol. 26, n°4, 2008, pp. 550-562.

[15] E. Trucco, A. Verri, "Introductory Techniques for 3-D Computer Vision", Prentice Hall, 1998.

[16] O. Faugeras, "Three Dimensional Computer Vision", MIT Press, 1993.

[17] R. Hartley, A. Zissermann, “Multiple view geometry in computer vision”, Cambridge university press, 2nd edition, 2003

[18] A. Del Bue and L. Agapito. “Non-rigid Stereo Factorization”, *International Journal of Computer Vision (IJCV)*, vol. 66, n°2, 2006, pp. 193-207.

[19] A. Del Bue, X. Llado and L. Agapito, “Non-rigid Face Modelling Using Shape Priors”, *Analysis and Modelling of Faces and Gestures (AMFG 2005)*. Springer LNCS 3723.

[20] A. Del Bue, L. Agapito, “Non-rigid 3D shape recovery using stereo factorization”. Asian Conference of Computer Vision (ACCV2004), 2004, vol. 1, pp. 25-30, 2004.

[21] R. Hartley, A. Zissermann, "Multiple view geometry in computer vision", Cambridge university press, 2nd edition, 2003
[22] N Ayache, P.T. Sander, "Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception" MIT Press, 1991.
[23] R. Hartley, "Stereo from uncalibrated cameras", Proc. Of the Second Joint European – US Workshop on Applications of Invariance in Computer Vision, 1994, pp. 237-256.
[24] Z. Zhang, "Determining the Epipolar Geometry and its Uncertainty: A Review", *International Journal of Computer Vision*, vol. 27, n°2, 1998, pp. 161-198.
[25] Point Grey Research. http://www.ptgrey.com
[26] S. Westland and A. Ripamonti, *Computational* Colour Science. John Wiley, 2004.
[27] A. Tremeau, C. Fernandez-Maloigne, and P. Bonton, *Image numérique couleur,* Dunod, 2004.
[28] M. J. Todd, and E. A. Yıldırım, "On Khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids", Discrete Applied Mathematics, vol. 155, Issue 13, 2007, pp. 1731-1744.

Obtaining Initial Controller Parameters for Approximate Pole Placement Iterative Feedback Tuning

J. Sikaundi
Department of Electrical Engineering
University of Cape Town
Private Bag, Rondebosch, 7701 South Africa

Abstract – **Changes are made to the approximate pole placement iterative feedback tuning algorithm, allowing it to be performed in a simpler manner. Two methods of finding the initial controller parameters for the approximate pole placement algorithm are presented and investigated. The first is based on virtual reference feedback tuning and the second on offline iterative feedback tuning. The initial controller parameters obtained are close to the optimum and thus allow the algorithm to settle in less iterations.**

I. Introduction

In closed loop design the poles of a process can either be shifted or cancelled depending on the design requirements of the engineer [1]. The disadvantage of cancelling the process poles is that they remain in the transfer function from the input disturbance to the output as shown in [2]. This may result in a process with poor internal performance [3]. Iterative feedback tuning (IFT) can at times result in a controller that cancels the process poles [4]. The reader is referred to [5] for an explanation of the IFT procedure. Four different methods for preventing pole-zero cancellation with IFT are also presented in [2]. The most useful of these methods is approximate pole placement iterative feedback tuning (APPIFT) since its resultant closed loop matches the user defined response in terms of damping factor and speed of response (for setpoint tracking and for disturbance rejection). However, a way of obtaining initial controller parameters was not suggested though good initial parameters will allow the optimum position to be reached rapidly when performing APPIFT. In addition it would prevent the algorithm from settling at local minimum points.

The recommended way of obtaining an initial controller for classical IFT is Virtual Reference Feedback Tuning (VRFT) [6, 7]. However it will be shown later that it cannot be applied for approximate pole placement. The VRFT technique can also produce a controller that cancels the process poles [8]. This paper presents a simpler manner of performing APPIFT to make it more useable by finding reasonable initial controller parameters. After the modification, two methods for obtaining initial controller parameters are presented. The first method is based on VRFT and the second on closed loop identification using IFT. Similar to IFT, both these methods operate in the closed loop. These new methods are simulated on different processes and then applied to a DC motor for speed control.

II. Modification of APPIFT

The basic idea of pole placement presented in [2] is that it places the controller zero in the desired closed loop model, and so the only difference between the closed loop and the desired model is the characteristic equation. When IFT is performed on this configuration the controller will use all its degrees of freedom on achieving pole placement. Given that the controller zero is always changing, the final achievable model is also unknown. Thus traditional VRFT cannot be performed for pole placement since it requires the inverse of the desired model. This motivates the research into finding a different approach for performing APPIFT.

A. The Development of the Algorithm

Consider the controller configuration shown in Fig. 1. Suppose that the feedback filter f(s) = 1 and that the prefilter p(s) was made a filter with poles at the position of the controller zero and a low frequency gain of one. Consequently the transfer function from the closed loop setpoint, *r,* to the output, *y*, will have no zeros in its numerator. Therefore APPIFT can be performed without inserting the controller zero into the desired model because the closed loop does not contain any zeros. Due to this, the controller zeros cannot cancel the process poles and the controller will use its degrees of freedom on pole placement. The reason is simply that the only difference between the closed loop and the model are the pole positions because the closed loop zeros have been cancelled by the definition of p(s). The only change from the normal IFT for a one degree of freedom control structure is the inclusion of a prefilter as defined above. Hence this prefilter is also updated at each IFT cycle. The IFT equations remain the same thus this method becomes a simpler way of implementing APPIFT because the model output derivative with respect to the controller parameters is now zero.

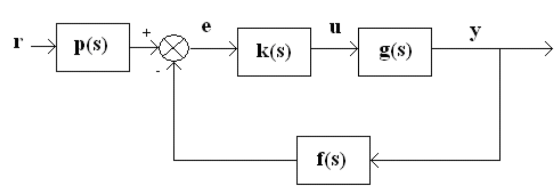


Fig. 1. Control loop under investigation

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_9,

These observations are now demonstrated in the mathematical equations below. For the loop considered, the transfer function for setpoint-tracking is:

$$h_{yr} = \frac{n_p}{d_p} \times \frac{n_k n_g}{n_k n_g + d_k d_g} \tag{1}$$

where each transfer function is expressed as a ratio of polynomials (e.g. $g(s) = \frac{n_g}{d_g}$). If the controller zero is cancelled by the prefilter pole, $d_p = n_k$ then (1) becomes:

$$h_{yr} = \frac{n_p n_g}{n_k n_g + d_k d_g} \tag{2}$$

The numerator n_p can then be varied to give the transfer function unity gain at low frequency for setpoint tracking.

The case study in this paper considers a closed loop configuration with the first order process $g(s) = \frac{A}{1+Ts}$ and a PI controller $k(s) = \frac{b_1 s + b_0}{s}$ since many industrial processes can be approximated as first order dynamics. The aim is to obtain initial controller parameters for the configuration investigated in [2].

B. Algorithm Applied on First Order Processes

When the procedure is applied to a first order process the closed loop transfer function becomes:

$$h_{yr} = \frac{\frac{Ab_0}{T}}{s^2 + \frac{(Ab_1+1)s}{T} + \frac{Ab_0}{T}} \tag{3}$$

For the above to be equal to the second order model:

$$m(s) = \frac{d}{s^2 + cs + d} \tag{4}$$

the controller parameters will be:

$$b_0 = \frac{dT}{A}$$
$$b_1 = \frac{cT-1}{A} \tag{5}$$

Therefore pole placement to $\varphi_c = s^2 + cs + d$ can be achieved.

This was simulated for a number of first order processes, where the aim is to place the closed loop poles at s = -1+/-j0.5. The initial controller parameters were $k(s) = \frac{1+1s}{s}$. The results are tabulated in table 1 and confirm that approximate pole placement had occurred for all processes.

TABLE 1

APPIFT USING p(s) APPLIED ON FIRST ORDER PROCESSES

g(s)	b_0	b_1	Closed loop h(s) poles
$\frac{1}{1+2s}$	2.525	3.043	-1.011+/-j0.491
$\frac{1}{1+3s}$	3.769	5.043	-1.007+/-j0.492
$\frac{1}{1+4s}$	5.021	7.047	-1.006+/-j0.493
$\frac{1}{1+5s}$	6.268	9.049	-1.005+/-j0.494
$\frac{1}{1+6s}$	7.519	11.052	-1.004+/-j0.495

The procedure can be extended to higher order processes by increasing the complexity of the controller. For example, doing approximate pole placement of second order processes would require a controller such as $k(s) = \frac{b_2 s^2 + b_1 s + b_0}{s(sT_c+1)}$ where a non-dominant filter is added for causality.

III. A MODIFICATION OF VRFT

The reader is referred to [6] for a thorough explanation of the VRFT technique which is summarised in this section for convenience.

The idea of VRFT is that there is a process output $y = \frac{kg}{1+kg} r$ in closed loop for a one degree of freedom control structure (i.e. p(s) = f(s) = 1 in Fig.1) and a desired model output, $y = mr$. When the closed loop and the desired model are the same then the input into the two transfer functions will give the same output, *y*. There is also a specific process output, *y* when a process has a particular input i.e. $y = gu$. Suppose that a known set of data of *u* is used as an input into the process *g* to obtain another set of data *y*. Using this particular output data set, *y*, the setpoint, $\underline{r}$, required by the desired closed loop model, *m*, to obtain an output equivalent to *y* can be calculated. This is done by passing the given output through the inverse model (i.e. $\underline{r} = m^{-1} y$). This setpoint, $\underline{r}$ is referred to as the virtual reference as it was not applied to the closed loop.

In closed loop the input into a controller, *k*, is the error, *e*, between the setpoint and output. This error, *e*, can be obtained using the virtual reference and the output i.e. $e = \underline{r} - y$. When both the closed loop and the desired model have this virtual reference as a setpoint, the outputs will be the same for the optimal controller. Thus the optimal controller is one that produces the vector of data *u* for the computed *e*.

As a result the criterion to be minimised is:

$$J = \frac{1}{N}\sum_{t=1}^{N}(u(t) - k(s;\rho)e(t))^2 \tag{6}$$

where ρ represents the parameters of the controller

Let $k(s;\rho) = \beta^T(s)\rho$. Then (6) becomes:

$$J = \frac{1}{N}\sum_{t=1}^{N}(u(t) - \varphi^T(t)\rho)^2 \tag{7}$$

where $\varphi(t) = \beta^T(s)e(t)$ is computed by digital simulation.

When the optimum parameters have been obtained, the cost function is zero and the parameter vector is defined by:

$$\rho_N = \left[\sum_{t=1}^{N}\varphi(t)\varphi(t)^T\right]^{-1}\left[\sum_{t=1}^{N}\varphi(t)u(t)\right] \tag{8}$$

VRFT settles at a global minimum but is only exact if the optimum controller belongs to the class of controllers being tuned. IFT optimizes its cost function locally [7, 9], and this makes the use of VRFT a complementary technique for IFT which is optimum for all controllers.

Since VRFT can be used to obtain the initial parameters that can be fine tuned by IFT, there needs to be a method to obtain the initial parameters for the APPIFT algorithm. VRFT uses the inverse of the model to obtain the optimal parameters needed to achieve the closed loop model. Unfortunately the model zeros are unknown in pole placement since these are the zeros of the optimal controller that is being sought. Hence, the traditional VRFT cannot be used for pole placement.

Given that the pole placement algorithm in the previous section performs pole placement using a prefilter that cancels the controller zeros a different approach can now be used. VRFT for pole placement is developed in this section and is referred to as approximate pole placement virtual reference feedback tuning (APPVRFT). Once again consider the case of a PI controller and a first order process. This gives the closed loop transfer function $h(s) = \dfrac{\frac{A}{T}(b_1 s + b_0)}{s^2 + \frac{Ab_1+1}{T}s + \frac{Ab_0}{T}}$. Clearly the zero positions that would achieve pole placement are not known because there is no process model. However if the zero of the controller is cancelled by a prefilter, p(s), on the setpoint the closed loop transfer function becomes $h(s) = \dfrac{\frac{Ab_0}{T}}{s^2 + \frac{Ab_1+1}{T}s + \frac{Ab_0}{T}}$ and therefore there are no zeros in the closed loop and so no zeros are needed in the model. Thus the model can be of the form $m(s) = \dfrac{d}{s^2 + cs + d}$ and the normal VRFT procedure can now be applied. The inverse of the above model will be needed; therefore non-dominant poles will have to be added to the inverse model $m(s)^{-1} = \dfrac{s^2 + cs + d}{(es^2 + fs + d)}$ where e and f will have to be chosen carefully. This will allow for the reference model to be obtained. The error signal is now needed to obtain the controller that will give that output when the input is the virtual reference signal $e_p = p\underline{r}-y$ in which p is not known since it is dependent on the final controller numerator, though it has a gain of one. The process input $u = ke_p = k(p\underline{r}-y)$ so that:

$$u = \frac{b_1 s + b_0}{s}\left(\frac{b_0}{b_1 s + b_0}\underline{r} - y\right)$$
$$= \begin{pmatrix}\dfrac{e}{s} & -y\end{pmatrix}\begin{pmatrix}b_0 \\ b_1\end{pmatrix} \tag{9}$$

Therefore $\varphi^T = \begin{pmatrix}\dfrac{e}{s} & -y\end{pmatrix}$ can be used in (8) to obtain:

$$\begin{pmatrix}b_0 \\ b_1\end{pmatrix} = \begin{bmatrix} \frac{1}{N}\sum_{t=1}^{N}\left(\frac{e_t}{s}\right)^2 & \frac{1}{N}\sum_{t=1}^{N}\left(\frac{e_t}{s}\right)\times -y_t \\ \frac{1}{N}\sum_{t=1}^{N}\left(\frac{e_t}{s}\right)\times -y_t & \frac{1}{N}\sum_{t=1}^{N}y_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{N}\sum_{t=1}^{N}\left(\frac{e_t}{s}\right)\times u_t \\ \frac{1}{N}\sum_{t=1}^{N} -y_t \times u_t \end{bmatrix} \tag{10}$$

The challenge is to select the position of the non-dominant poles of the inverse model to obtain the virtual reference. This will also pose a restriction on the sampling time for digital implementations.

Since the values of the input u and the output y are known, the equivalent z-transform of the model can be used to calculate the reference signal parameters at each sampling time. From this vector of reference signals, $\underline{r}$, the value of parameters b_0 and b_1 can be calculated.

As before, pole placement to the pole positions s = -1+/-j0.5 is desired. For a sampling time of 0.02, this will give an equivalent z-transform of the form $m(z) = \dfrac{0.00049}{z^2 - 1.9604z + 0.96089}$ for the given m(s). Since the y values are known the $\underline{r}$ values at each sample time can be calculated using the equation below.

$$m(z)^{-1} = \frac{r}{y} = \frac{1 - 1.9604z^{-1} + 0.96089z^{-2}}{0.00049z^{-2}}$$
$$\Rightarrow r_t = \frac{y_{t+2} - 1.9604y_{t+1} + 0.96089y_t}{0.00049} \tag{11}$$

From this, the error e can be calculated and hence the optimum parameters b_0 and b_1. Table 2 shows the results of

TABLE 2

APPVRFT APPLIED ON FIRST ORDER PROCESSES

g(s)	b_0	b_1	Closed loop h(s) poles
$\frac{1}{1+2s}$	2.463	2.990	-0.998+/-j0.486
$\frac{1}{1+3s}$	3.688	4.975	-0.996+/-j0.488
$\frac{1}{1+4s}$	4.913	6.960	-0.995+/-j0.488
$\frac{1}{1+5s}$	6.139	8.945	-0.994+/-j0.489
$\frac{1}{1+6s}$	7.394	10.930	-0.994+/-j0.489

applying this method on various first order processes when using the initial controller $k(s)=\frac{1+1s}{s}$.

It can be seen in Fig. 2 that the controller parameters obtained by APPVRFT are not dependent on the initial damping factor of the closed loop. However there is a difference between the optimum controller parameters and those obtained using APPVRFT. It was found that these differences decreased as the sampling time was decreased. Since the sampling time is limited in physical implementations, APPIFT can be used for fine tuning.

As a one shot method APPVRFT does not produce a controller that cancels the process poles and it has the advantage of normal VRFT that it settles at a global minimum point.

IV. OFFLINE POLE PLACEMENT IFT

Normal IFT is done by tuning the controller parameters so that the closed loop can resemble that of the desired model in a least squares sense according to a cost function such as the errors squared cost function shown below.

$$J=\frac{1}{2N}\left[\sum_{i=1}^{N}(y-ym)^2\right] \tag{12}$$

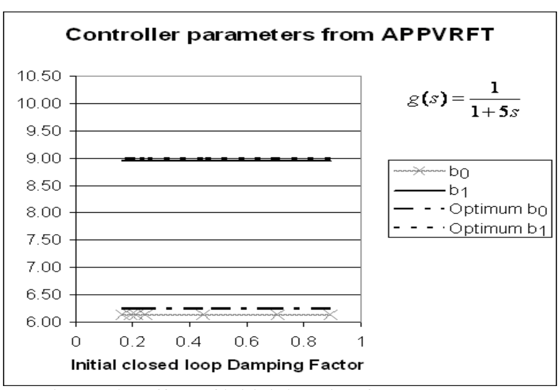


Fig. 2. The effect of initial damping factor on APPVRFT

The derivative of the cost function with respect to the controller parameters is used to tune these parameters iteratively. The model output derivative with respect to the controller parameters is zero so the model parameters remain constant.

Consider again the case of a PI controller and a first order process. As shown earlier, another manner of performing pole placement is to cancel the controller zero from the closed loop using a prefilter and a model with no zero of the form: $m(s)=\frac{d}{s^2+cs+d}$. The closed loop transfer function for any controller parameters will be $h(s)=\frac{\frac{Ab_0}{T}}{s^2+\frac{Ab_1+1}{T}s+\frac{Ab_0}{T}}$ because the prefilter has cancelled the controller zeros.

In this section a different manner of performing pole placement is suggested. IFT should be performed and optimised according to what model parameters would result in the smallest cost function in a least squares sense. Therefore the first IFT experiment is performed by applying a setpoint to the closed loop and the chosen initial model. During this experiment the output, y and the model output ym are collected. The difference is that the first IFT experiment happens once on the closed loop. The data is stored and continuously used to tune the model so that its response to a setpoint resembles the closed loop output in the first IFT experiment. Thus for the cost function stated earlier the derivatives with respect to the model parameters, σ:

$$\frac{\partial J}{\partial \sigma}=-\frac{1}{N}\sum_{i=1}^{N}em\frac{\partial ym}{\partial \sigma} \tag{13}$$

are needed. The model is:

$$m(s)=\frac{d_0}{s^2+d_1s+d_0} \tag{14}$$

where d_n are variables depending on the process g(s).

The model can then be tuned offline using the same IFT algorithm i.e. $\sigma_{i+1}=\sigma_i-\gamma_i R_i^{-1}\frac{\partial J}{\partial \sigma}(\sigma_i)$.

Once the optimal parameters have been obtained (for d_0 and d_1) the controller parameters that can give pole placement to the desired location can be obtained. Therefore since d_0, d_1, b_0 and b_1 are known, the values of A and T can be calculated. Using these values of A and T the required values of $\underline{b_0}$ and $\underline{b_1}$ that would give the desired pole positions can be calculated:

$$\underline{b_1}=\frac{cT-1}{A}$$

$$\underline{b_0}=\frac{dT}{A} \tag{15}$$

Therefore this method uses iterative identification to perform pole placement and will be referred to as approximate pole placement iterative feedback tuning modelling (APPIFTM). Normal APPIFT can then also be used to tune the system so that the bias of noise in the original signal is eliminated. The results of applying this method to a number of first order processes are shown in Table 3.

An advantage of APPIFTM is that if used iteratively, the user will know the transfer function of the closed loop and can make sure that the closed loop, h(s) has the dynamics of the model, m(s). Another advantage is that if the closed loop cannot be modelled by a second order model then the order of the process is not first order. From this the user can determine that pole placement for a second order processes should be attempted.

However the disadvantage is that the user needs to know the form of the model of the closed loop, and this requires that the user should be able to estimate the approximate order of the process dynamics. Another disadvantage is that it also requires an initial closed loop model before beginning IFT. It is recommended that this initial model be the desired model to avoid having to choose a new model. A further disadvantage is that since this method uses the normal IFT procedure, a gradient descent algorithm, it could become trapped in a local minimum.

The effect of the initial closed loop damping factor on the obtained parameters is shown in Fig. 3. Unlike APPVRFT the controller parameters are dependent on the initial damping factor of the closed loop.

V. Application On a DC Motor

When APPIFT was applied to a motor system for speed control the results were as shown in Fig. 4 and Fig. 5. The aim was to place the closed loop poles at s = -0.4+/-j0.2 and the programming sampling time was 250ms. The initial controller is $k(s) = \frac{1s+1}{s}$ and the nominal final controller after 250 IFT iterations was $k(s) = \frac{1.770s + 0.526}{s}$.

TABLE 3

APPIFTM Applied On First Order Processes

g(s)	b_0	b_1	Closed loop h(s) poles
$\frac{1}{1+2s}$	2.564	3.111	-1.028+/-j0.475
$\frac{1}{1+3s}$	3.816	5.116	-1.019+/-j0.483
$\frac{1}{1+4s}$	5.039	7.094	-1.012+/-j0.486
$\frac{1}{1+5s}$	6.397	9.200	-1.020+/-j0.489
$\frac{1}{1+6s}$	7.543	11.097	-1.008+/-j0.491

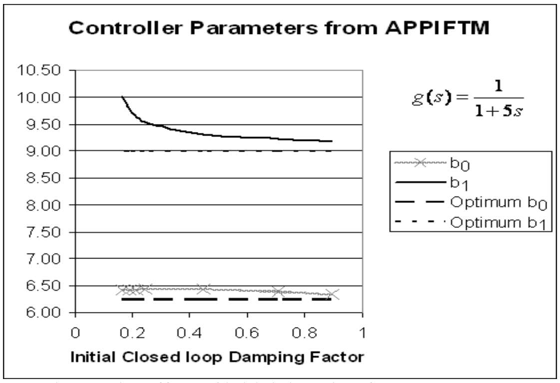


Fig. 3. The effect of initial damping factor on APPIFTM

The nominal transfer function for the DC motor was found to be $g(s) = \frac{3.0 \pm 0.1}{1+(7.5 \pm 0.3)s}$ through step tests on the open loop. The chosen operating region was 5 - 8 volts to avoid the non-linear dead-band region of the DC motor around 0 volts and the maximum voltage reading of the ADC of 10 volts.

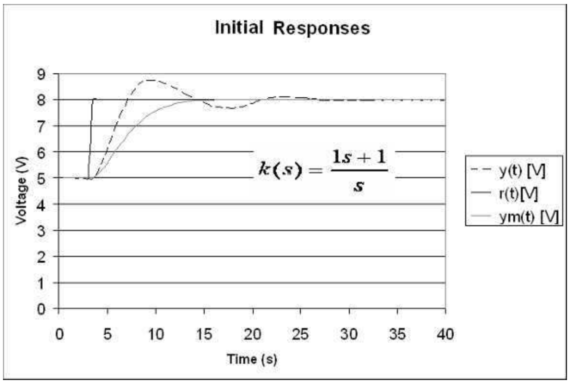


Fig. 4. Responses with initial controller

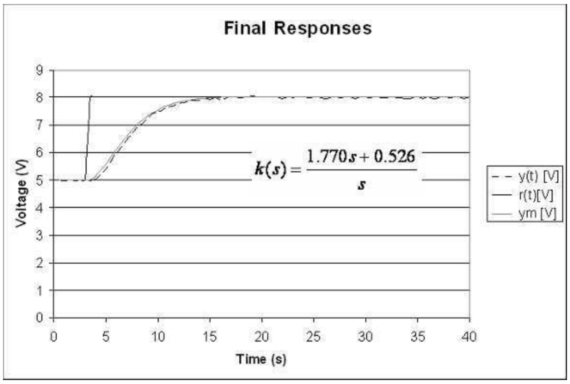


Fig. 5. APPIFT Responses with final controller

When the APPVRFT was applied on the closed loop with the same initial controller as before and the equivalent desired model $m(z) = \dfrac{0.000113}{z^2 - 1.8074z + 0.8187}$, the final response is that shown in Fig. 6. The response of the process after one application of APPIFTM is also shown in Fig. 7.

Judging from the responses in Fig. 6 and Fig. 7, it appears that both methods give reasonable initial controller parameters. However APPVRFT appears to give parameters that are closer to the optimum, $k(s) = \dfrac{1.770 + 0.526}{s}$, obtained using APPIFT. Its advantage is that it does not use an intermediate stage of identification before operation. The advantage of the APPIFTM on the other hand is that it can be applied iteratively until the closed loop directly matches the model and the user will know the transfer function of the closed loop. Its algorithm is also the same as that of IFT.

VI. CONCLUSION

The APPIFT method is considered the best way of preventing pole-zero cancellation. However it previously suffered from not having an automatic way of finding initial controller parameters that are close to the optimum parameters. Two novel methods were presented for finding initial controller parameters for this technique. Both are one shot methods and give a reasonable starting closed loop with a response close to the optimum desired performance. This will allow the APPIFT method to be used for fine tuning if the closed loop obtained is not satisfactory.

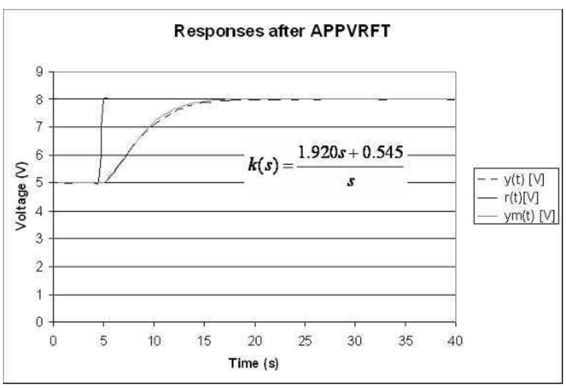


Fig. 6. Responses after performing APPVRFT

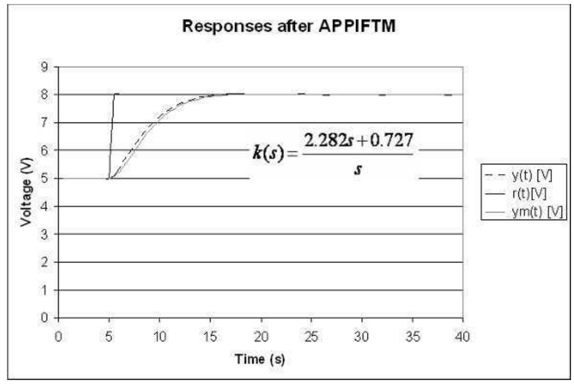


Fig. 7. Responses after performing APPIFTM

ACKNOWLEDGMENT

The financial assistance of the Department of Labour (DST) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the DST.

REFERENCES

[1] R.H. Middleton and S.F. Graebe, "Slow stable open-loop poles: to cancel or not to cancel," *Automatica*, vol. 35, pp. 877-886, 1999.

[2] J. Sikaundi and M. Braae, "Preventing pole-zero cancellation for Improved input disturbance rejection," *Proc. CISSE Int. Joint Conf.* pp. 111-116, December 2007.

[3] P.A. Laplant, "Comprehensive dictionary of electrical engineering," CRC IEEE Press, 1999.

[4] H-Y. Wei, S.I. Ginsberg and M. Braae, "Engineering an iterative feedback controller". *Proc.SACAC Control Conf*, pp. 64-69, July 2006.

[5] H. Hjalmarsson, M. Gevers, S. Gunnarson and 0. Lequin, "Iterative feedback tuning: theory and applications," *IEEE Control Systems*, pp. 26-41, August 1998.

[6] M.C. Campi, A. Lecchini and S.M. Savaresi, "Virtual reference feedback tuning: a direct method for the design of feedback controllers," *Automatica*, vol. 38, pp. 1337-1346, 2002.

[7] G.E Solari, "Iterative model-free controller tuning," Ph.D. Thesis. Universite Catholique de Louvain. 2005 [Online]. Available: http://edoc.bib.ucl.ac.be:81/ETD-db/collection/available/BelnUcetd-08042005-092900/unrestricted/IFT_Thesis.pdf [2007, April 22]

[8] A. Lecchini, M.C. Campi and S.M Savaresi, "Virtual reference feedback tuning for two degree of freedom controllers," *Int. J. Adapt. Control Signal Process, special issue on controller design*, vol. 16, pp. 355-371, 2002

[9] R. Hildebrand, A. Lecchini, G. Solari and M. Gevers, "Prefiltering in iterative feedback tuning: optimization of the prefilter for accuracy," *IEEE Transactions on Automatic Control*, vol. 49, pp. 1801-1805, October 2004.

[10] A. Lecchini and M. Gevers, "On iterative feedback tuning for non-minimum phase plants," *Proceedings of the 41st IEEE Conference on Decision and Control Las Vegas*. Vol. 4, pp. 5658-4663, December 2002.

Organizational Learning and Knowledge Management

Dr. PhD Margareth Stoll
Free University of Bozen
Sernesistraße, 1
Bozen, South Tyrol, Italy
Margareth.stoll@dnet.it

Abstract **Standard based management systems (as quality ISO9001, environmental ISO14001, information security ISO27001, hygiene management systems ISO 22000 and others) are widely implemented and based on common principles: objectives and strategies, business processes, resource management and continuously optimization. These systems contains all relevant information, regulations, forms and process descriptions and must also be documented (system documentation), communicated, implemented and continuously improved. Therefore they represent an optimal basis for a systematic and well structured management of regulations, if the requirements of management systems are completely and the organization adapted implemented. Although this documentation is distributed recently usually IT supported it is almost felt as additional workload with a little or no advantage, it is not totally corresponding with the lived processes. Therefore it is hardly used as reference book for solving job problems and questions. It is not basis for a structured, systematic regulation management to improve the organization.**

Based on this situation we prepared the regulations and system documentation according to media-pedagogical and didactical principles and published it on organizational document learning and knowledge system based on constructivist theory. In the case study the regulations promotes in a confidence-based, open and fault-tolerant corporate and learning culture a need-oriented workplace process integrated regulation access, organizational development and knowledge generation, employee involvement and knowledge sharing, practice-oriented regulation and system documentation, shorter initial training periods for new collaborators and a continuous optimization of the organization for securing a sustainable organization success.

Keywords: regulations, document management, standard based management systems, need oriented access, knowledge management, optimization of organization

I. Introduction

A Starting Situation

By globalization and ever shorter change cycle's organizations must improve increasingly faster their products, services, technologies and organization according to customer requirements. Thus the largest potential is the continually improvement, the management of information, documents and knowledge, the sustainable continuous need-oriented, organizational learning due to individual learning and in a holistic interdisciplinary development approach. For an efficient, effective and sustainable organizational development are more frequently integrated different relevant aspects into a holistic management system. More than one million organizations of different sizes and scopes are implementing, already since several years, management systems, such as ISO9001, ISO14001, ISO27001, ISO22000 or others [1], which are based on international standards with common principles: organization objectives and strategies, business processes, resource management and continuously optimization of the organization. The established management system must be documented, communicated, implemented and continuously improved. Thereby the regulation management must be implemented in a structured and systematic way. Therefore the system documentation should contain all up to date and necessary information, documents (organization regulations, forms, process description) and thus contain the entire clearly structured explicit organizational knowledge. Individual learning and thus organizational learning and the sustainable organizational development are promoted by an effective, need-oriented, process integrated access to regulations.

B Purpose of the article

The documentation of standard based management systems (system documentation) was distributed for long time as books, whereby the collaborators received one-time the information and used them afterwards for possible questions scarcely ever. In the last ten years the system documentation was distributed more electronically through web-based intranets, document management platforms or as pdf. The delivered system documentation is in theses way used a little more as reference book. But changes proposal are not related to the documentation and the collaborators are not able to discuss new ideas or questions in a context-sensitive way. The organizational improvement is isolated from the system documentation. The collaborators read it once when it is distributed, but is not used for solving ad hoc job problems. It is usually changed only once a year, if at all and it frequently does not correspond to lived processes. Moreover in addition

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_10,

to it instructions and forms are developed as hidden system and these call the access and integrity of the whole system documentation in question.

C *Research approach and structure of the article*

The requests for always shorter innovation cycles, the information explosion, the increasing rules and regulations, the ever stronger IT involvement and the great importance of organizational development, the continually organizational and individual learning for the sustainable organization success and the central role regulations and information leaded us to the extension of a constructivist learning platform as an organizational learning and knowledge system [2], [3] for distributing the system documentation of a standard based and knowledge integrated management in accordance to media-pedagogical and didactical principles and introduced it in a confidence-based, open and fault-tolerant corporate and learning culture.

Firstly we illustrate the main requirements of international standards for management systems [II] and based on the starting situation we describe the project objectives [III]. Afterwards we explain our approach [IV] including the requirements for an organizational learning and knowledge system [IV A] and the success factors [IV B]. Finally we document the project experiences and results of the implementation in different organizations with distinct management systems [V] including the achievement of the project principles [V A] and a reflection about cost and benefits [V B] and at the end we express an outlook [VI] and our conclusion [VII].

II. Main Requirements of Standard Based Management Systems

The standards for management systems have different specialized focuses, but are all based on common principles [4], [5]:

- Organizations objectives and strategies must be established regarding customer requirements.
- All business processes including management process, support processes, resource processes and optimization processes must be defined and promote the optimized fulfillment of the organization objectives under the focus of the respective standard.
- Process oriented resource management must be promoted including personnel development, IT – management and other infrastructures, tools and instruments.

Fig. 1. Main requirements of standard based management systems

- The organization and their objectives and strategies must be continually optimized according to established processes in sense of a PDCA cycle (plan, do, check, act).

The established management system must be structured and systematically documented and communicated within the organization and the collaborators must be continually motivated for implementing the system and for recommending improvements.

III. Project Objectives

How can information, regulations, documents (e.g. forms, process description, checklist) prepared systematically and stored well structured to promote need-oriented access for the collaborators to get there job done efficiently and effectively to bring advantages for the organization and how can they be constantly developed?

By preparing the documents and regulations of a holistic, standard based, the organization adapted management system in accordance to media-pedagogic and didactical principles systematically and structured and implementing it on an organizational learning and knowledge system by extending a constructivist learning system we expect to foster:

- need-oriented workplace process integrated regulations access
- organizational development and knowledge generation,
- employee involvement and knowledge sharing,
- practice-oriented regulation and system documentation,
- shorter initial training periods for new collaborators.

Thus the structuring of the regulations and documents, need-oriented workplace process integrated access to get the job done efficiently and effectively, learning, knowledge generation, knowledge representation, knowledge communication and the application of the regulations should be optimized and the organization will be promoted in achieving its objectives including knowledge objectives.

IV. Approach

Firstly the system documentation must be established. Considering the needs and expectations of all interested parties we establish the organization policy with consistent strategies and goals integrating knowledge objectives and plans [6], [7]. Afterwards the applied processes are analyzed bottom up by interviewing the collaborators involved and by integrating knowledge management and optimized in accordance with the objectives of the organization. The applied process modeling method defines regarding the organizational objectives and the legal and normative regulations for all activities the responsible function, the necessary documents, information, tools and it – applications. Thereby all responsible for the creation, change, change, elaboration of the documents and data, all needed access rights, the external document and data sources and receiver, further the archiving procedures including duration and necessary information security regarding document logistic and data protection are determined.

The necessary resources, tools, instruments, trainings required for achieving the objectives and for improving the management system must be identified, made available and continually optimized. Also the training and human resource development processes must be optimized and established. The organization must also plan and implement the monitoring, measurement, analysis and optimization processes needed to continually improve the effectiveness of the management system. Thus also knowledge management and learning organization must be integrated and their improvement must be structured and systematically planned [8], [9], [10].

The entire system documentation must correspond with the lived processes [11]. All improvements or changes must be approved by the involved collaborators, documented, communicated and implemented.

After the development of the management system documentation, it must be prepared regarding media-pedagogical, motivation-psychological and didactical principles [12]. According to these principles it must be divided into small modules, which must be typecasted and functionally well structured. Thus, we use a constructivist method for organizing the content. In order to support the collaborators need-oriented workplace integrated access, the way for accessing the single modules should be as short as possible, optimal structured and practice oriented. All needs and expectations of the collaborators must be analyzed and considered as early as possible [11], [13]. They demand particularly an efficient and effective search function and a clear structured system. Therefore we need effective indexing and different start assistances (for new collaborators, department oriented, management oriented, topics referred, based on the standard and others). Apart of the self-driven learning approach, the system must also offer guided learning for new collaborators or for collaborators with little IT competences.

After the appropriate editing the system documentation can be published on the organizational knowledge and learning system. The upload of the content should be user-friendly and simply, because it is normally an activity done by the system manager (responsible for implementing and supporting the management system), which has organizational, managerial and controlling competences and competences in developing management systems and must not be an IT expert. The constructivist approach in learning requires the possibility of linking concepts and objects following one's own needs. This feature must be available also for graphical objects and different document and media formats, which are widely used in hypertext contents.

To promotes the acceptance of the system the collaborators must be trained on the handling the system and acquire also the necessary media competence [11]. Subsequently, questions will be answered by means of the discussion forum, the help desk and personal partners, e.g. the knowledge and/or system manager. Using the constructivist organizational learning and knowledge system every collaborator can introduce his/her suggestions, ideas or questions directly in a context-sensitive way by referencing all the possible objects of the system documentation. Thus, ambiguous formulations in the documentation are showed up and the system documentation - organizational knowledge base - becomes the basis for the continuously optimization of the organization. Thereby we integrate also optimal process modeling, process standardization and transparency with need and objective oriented flexible process implementation.

The collaborators have sometimes problems to distinguish between problem and optimization, or they do not know to which section of the system documentation they should relate their suggestion or question. Therefore we have integrated the organizational knowledge and learning system with the existing workflow driven problem reporting system (help desk system).

Following the management system requirements also clearly structured processes for the introduction of new collaborators, for the maintenance of the system documentation and for the management of questions, optimizations and problems must to be established, documented, implemented and maintained.

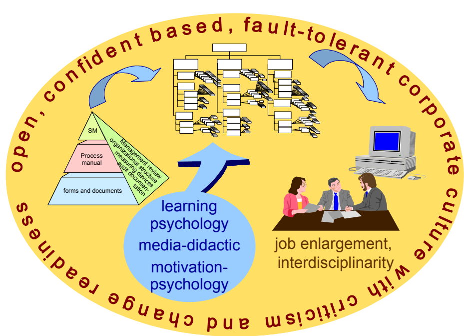


Fig. 2. Approach for need oriented regulations access

A *Requirements for an organizational learning and knowledge system*

Based on our project experiences an organizational learning and knowledge system for standard based management systems demands, in addition to the general requirements of learning or knowledge management systems, the following particular characteristics:

- It must be simple and intuitive to handle, it must provide the possibility to use different views of (or leading to) the same object, different start facilities, comfortable search functions, filtering of content using object types, simple uploads of content and links to external literature and it must promotes individual learning and organization by allowing personal bookmarks, annotations, summaries and notes, as well as glossary, FAQs, etc.. [13], [14], [15].
- It must offer secure context-sensitive communication to all elements (discussion forum, chat), as well as newsgroups, wiki, newsletters and whiteboards, support the knowledge distribution, the collective knowledge generation and collective learning [14], [16].
- Uploading and the administration of new content must be simple, with as large didactical and media pedagogic support as possible [2]. There must be the possibility to insert, annotate and discuss context sensitive the content of most different document and media formats (as text, graphic, table, picture, sound records, video). Also creativity tools should be integrated.
- Due to the requirements of the standards for management systems, we need the administration of access rights [17], the support of change processes, versioning with change history and the efficiently and traceable distribution / communication of new or changed content. Depending on the organization culture also testing and examination tools for the traceable learning must be integrated.
- The ideas of the collaborators, their discussion contributions and problem reporting must be implemented in accordance with the established systematic and structured process following the standard requirements and their effectiveness must be evaluated.
- Open interfaces must be available for fostering optimal connectivity with other systems in order to support process integration and simplify the administration [15], [17]. Within the controlling process all measured values should be evaluated by the responsible departments and consequently changes and optimizations should be implemented using the organizational learning and knowledge system.
- The web-based accessibility of the organizational learning and knowledge system supports mobile and working hour's independent learning and teleworking, making also easier the integration of absent collaborators [18].

B *Success factors for an organizational learning and knowledge system*

Corporate culture processes and IT technologies must be integrated optimally according to the objectives of the organization, to the collaborators needs and requirements and considering media-pedagogical and didactical principles, in order to successfully introduce an organizational learning and knowledge system into management systems. The system is thereby only a tool, which supports the optimization of the organization so far as this is admitted by the culture. Therefore we need an open, confident based, fault-tolerant corporate and learning culture with criticism and change readiness. The collaborators should be interested in new knowledge, able for self-driven learning, have personal employment, team ability and change willingness apart from necessary IT-competences. All managers must use constantly and actively the system and motivate their collaborators in following these principles, promoting in this way learning and new knowledge generation. A strengthening point for achieving this is maintaining an optimal communication internally as well as with external partners.

By introducing this system the system manager and/or knowledge manager extend their own job, needing also the acquisition of the necessary media-pedagogical and didactical knowledge for preparing the content and the necessary skills for supporting e-learning and knowledge management.

Sufficient IT-infrastructure and IT-support are also very important for the project success. Only by using an organizational learning and knowledge system, which meets as far as possible the stated requirements [IV A] and by promoting a workplace need-oriented process integrated learning, a continuously optimization of the organization, in accordance with its objectives, and a sustainable organization success are secured.

V. PROJECT EXPERIENCE AND RESULTS

The concepts expressed so far have been implemented successfully in several organizations with different management systems.

The interdisciplinarity was a great challenge and also the great chance. Knowledge about organization theory, management methods and standards were required for the preparing and structuring the content, information-technical knowledge were needed for extending the learning platform to meet the requirements of an organizational learning and knowledge system; didactical, media-pedagogical and communication science knowledge were required for editing the contents.

A *Achieving the project objectives*

The described concept has been implemented in medium-size service organizations by integrating knowledge management into management systems. Most of the organization's collaborators own good media competences and use frequently e-tools.
Nevertheless the platform used in the case studies fulfills only some of the system requirements [IV A] to an organizational learning and knowledge system], the adoption of the organization adapted system documentation according to media-pedagogical and didactical principles and taking into consideration a confident based fault-tolerant corporate culture, leads the following case study results collected by measuring the system accesses / user contributions and interviewing the leadership and collaborators:

- Need-oriented workplace process integrated regulations access: the accesses to the regulations are increased monthly at averaged two accesses per collaborator.
- Organizational development and knowledge generation: we receive five times more suggestions and ideas, which improves the organization.
- Employee involvement and knowledge sharing: the communicated ideas, problems and suggestions are discussed and read on the average by a quarter of the collaborators. Thus, the advantages and disadvantages of ideas are discussed and examined by all departments before their possible implementation and they are substantially more balanced and more considered for implementation.
- Practice-oriented regulation and system documentation: while conventional system documentation is usually adapted only once a year, now unclear formulations or missing regulations are soon analyzed and immediately changed according to established processes and automatically communicated to all. Thus, the documentation is adapted optimally to the changing requirements of the organization.
- Shorter initial training periods for new collaborators: new collaborators are quickly introduced into the system at their first working day and afterwards they learn in a self-driven need oriented way, all relevant contents for their specific work. Thus the documentation is always accessible the new collaborator can focus on the principle information and can after need-oriented access to all necessary regulations and process to fulfill his job. The lead time could be abbreviated around a quarter. New collaborators identify themselves substantially earlier with the corporate culture, they feel themselves also more integrated into the organization, can execute their tasks faster well, whereby the productivity increase and possible errors are avoided.

The manager of an organization with an organizational knowledge and learning system is particularly enthusiastic for the sustainable promotion of an open, confident based, fault-tolerant corporate and learning culture with criticism and change readiness.

The collaborators and particularly the management appreciate the structured, effective, need oriented, location and time independent access to regulations. Thus no unnecessary regulations are distributed and nevertheless all collaborators can access exactly at the appropriate time from the desired location to the necessary regulation. The improved internal transparency and the discussion board promote organizational interrelationship and synergies. The discussion of all suggestions is appreciated.

Standard based holistic management systems promotes by there clear structure and systematic document management (establish, check, approve, distribute and implement), the regulation structure, regulation retrieval and a process oriented document management.

Based on the experiences in different case studies the application of this integrated organizational learning and knowledge concept is recommended for medium and large sized organizations and enterprises with an confidence-based, open, innovative and fault-tolerant corporate and learning culture, where all collaborators command sufficient it competences, particularly in know-how enterprises, distributed locations, many collaborators in field service or with flexible working-time model.

B *Cost and Benefits*

This new interdisciplinary holistic organizational learning and knowledge concept requires in comparison to previously applied introduction of a standard-based holistic interdisciplinary the organization adapted system additionally the restructuring of the system documentation in accordance to media-pedagogical and didactical principles, an organizational learning and knowledge system and the implementation of the documentation on it. The restructuring required in the case studies a slightly higher effort (approximately 1-2% of the total effort).
Opposite to the costs are the large competition advantages through need-oriented process integrated regulation access to improve the organization by means of collaborators ideas, suggestions and discussion contributions. These advantages can be measured on one hand by the reached objectives of the project [5.1]. The influence to the sustainable achievement of the organizations objectives must be weighted essentially more strongly. In the sense of a holistic, systemic management the different strategies and there measurements influences each other and all individual actions together determine the achievement of the organizational objectives and the organizational success, which is at the end the essential critical factor. These effects must be evaluated in a longer term and could be objective of a subsequent study.

VI. OUTLOOK

Learning or knowledge management systems must be extended [IV A] to organizational knowledge and learning systems for need-oriented, workplace process regulations access.

By introducing an organizational knowledge and learning systems for standard based management systems with integrated knowledge management all success factors should be considered [IV B].

Training in system management, organization, knowledge management, information, document or regulation management or E-Learning consider more this interdisciplinary holistic and teach basic knowledge in all areas.

The standards for management systems should underline more and promote the importance of an open, confident based, fault-tolerant corporate and learning culture, knowledge management and integrated learning.

VII. CONCLUSION

Structuring all relevant information, regulations, forms and process descriptions of an organization regarding a integrated, holistic, standard based management system, preparing the system documentation in accordance to media-pedagogical and didactical principles and implementing it on a constructivist organizational learning and knowledge system with an open, confident based, fault-tolerant corporate and learning culture promotes need-oriented workplace process integrated regulations access, organizational development and knowledge generation, employee involvement and knowledge sharing, practice-oriented regulation and system documentation, shorter initial training periods for new collaborators and a continuous optimization of the organization by means of collaborators ideas and discussion contributions. Thus by fostering workplace need-oriented regulations access the sustainable success becomes guaranteed.

REFERENCES

[1] ISO, *ISO Survey of Certifications – 2006*, www.iso.org/iso/pressrelease.htm?refid=Ref1089, ISO, 2007..

[2] A. Back and O. Bendel Lern- und Wissensportale. in P. Glotz and S. Seufert Eds. *Corporate University*. Institut für Medien- und Kommunikationsmanagement, Universität St. Gallen, 2002.

[3] M. Bick and J.M. Pawlowsky A Framework for Integrated Ambient Learning and Knowledge Environments. in N. Gronau Eds. Proc. *4. Conference professional knowledge management* – experiences and visions. Gito, Berlin, volume 2, pp. 335342, 2007.

[4] ISO, *EN/ISO 9001:2000* Quality Managemnt Systems – requirements. ISO 17.12.2000.

[5] P. Osanna, M. Durakbasa and A. Afjehi-Sada, *Quality in Industry*, Vienna University of Technology, 2004.

[6] M. Stoll *Wissensmanagement in KMU durch die Erweiterung normbasierter Managementsysteme* in N. Gronau Eds. Proc. 4th Conference on Professional Knowledge Management – Experiences and Visions, Gito, Berlin (2007), volume 1 pp. 105-113.

[7] M. Stoll *Managementsysteme und Prozessorientiertes Wissensmanagement* in Proc. 4th Conference on Professional Knowledge Management – Experiences and Visions, Gito Verlag, Berlin (2007), volume 1, pp. 433-434.

[8] F. Lehner *Wissensmanagement*: Grundlagen Methoden und technische Unterstützung. Hanser, München, 2006.

[9] R. Maier *Knowledge management systems* - information and communication technologies for knowledge management. Springer, Berlin, 2002.

[10] G. Probst, S. Raub, K. Romhardt *Wissen managen*. Gabler, Wiesbaden, 1999.

[11] C. Bals, St. Smolnik and G. Riempp *A Case for Integrated Knowledge Management*. in N. Gronau Eds. Proc. 4. Conference professional knowledge management – experiences and visions. Gito, Berlin, volume 2, pp. 365–373, 2007.

[12] G. Reinmann-Rothmeier *Mediendidaktik und Wissensmanagement*. in MedienPädagogik 10, www.medienpaed.com/02-2/reinmann1.pdf, 2002.

[13] S. Teschler, C. Stracke and J.M. Pawlowski *Qualitätsmanagement in der E-Learning gestützten Aus- und Weiterbildung* – Anforderungen an ein Tool auf Basis einer Expertenbefragung. in M. Breitner, B. Bruns and F. Lehner Eds. Neue Trends im E-Learning. Physika, Heidelberg, 2007.

[14] A. Auinger and C. Stary *Didaktikgeleiteter Wissenstransfer*. Deutscher Universitäts-Verlag, Wiesbaden, 2005.

[15] G. Riempp *Integrierte Wissensmanagementsysteme*: Architektur und praktische Anwendung. Springer, Berlin, 2004.

[16] S. Smolnik and G. Riempp *Nutzenpotenziale Erfolgsfaktoren und Leistungsindikatoren von Social Software für das organisationale Wissensmanagement*. in HMD - Praxis der Wirtschaftsinformatik 252: pp. 17-26, 2006.

[17] B. Hildebrandt, C. Stracke, J.M. Pawlowski *Support Systeme für Qualitätsmanagement im E-Learning*. in M. Breitner, B. Bruns, F. Lehner Eds. Neue Trends im E-Learning. Physika, Heidelberg, 2007.

[18] K. Wilbers *E-Learning didaktisch gestalten*. in A. Hohenstein and K. Wilbers Eds. Handbuch E-Learning. Deutscher Wirtschaftsdienst, Köln, 2002.

System to Produce Mechanical Inertial Force and/or Torque

Dan Ciulin
E-I-A
Av. du Tribunal Fédéral 31, CH-1005, Lausanne, Switzerland.

Abstract - The Top and/or the gyroscope are devices able to produce inertial torques that maintains their direction in space if no external torques are applied. The resulted inertial torques are based on Coriolis forces. Two gyroscopes, placed as near as possible on a platform, may be used as a 'doublet' to produce inertial forces and/or torques. As the gyroscope is a 'gyrator', the efficiency of such a system is good. To be used as an inertial motor/torque for a vehicle, this 'doublet' system needs also an (electronic) processor to control it, a sensor for the position and direction in space and an supply source. For the security of the whole system, a given vehicle must be equipped with at least 2 similar 'doublets'. Then, such a system may be used as inertial motor/brake for terrestrial, submarine and/or even deep space ships. The big advantages of such a system are the facts that it needs not to eject masses as for the rocket engines, may be built inside the cell without any extensions to the exterior of the vehicle and may delivers inertial forces/torques big enough to vertically take off a given vehicle .

I. INTRODUCTION

Mechanical force $\vec{F}$ can appear due to the gravitational attraction between 2 masses $M_1.M_2$. That is mathematically modeled by:

$$\vec{F} = \frac{\gamma.M_1.M_2}{r^2}.\frac{\vec{r}}{|\vec{r}|}$$

(1)

where $\vec{r}$ is the position vector and γ the constant of universal gravitation ($\gamma \cong 6.6732.10^{-11} N.m^2.kg^{-2}$). This relation may be put also as:

$$\vec{F} = M_1.\vec{G}; \vec{F} = M_1.\vec{a}; \vec{G} = \vec{a} = \frac{M_2}{\gamma.r^2}.\frac{\vec{r}}{|\vec{r}|};$$

(2)

where $\vec{G}$ may be considered as a gravitational field and $\vec{a}$ as an acceleration. This realizes the connection between a 'static' gravitational field and a 'dynamic' moving mass. The acceleration can be in 'pure translation' and/or generalized, in translation and/or rotation. In this last case, some new accelerations can be discerned as centripetal $\vec{F}_{cen}$ and Coriolis $\vec{F}_{corr}$:

$$\begin{cases} \vec{F}_{cen} = M.\vec{\omega}\times\vec{\omega}\times\vec{r}; \\ \vec{F}_{corr} = 2.M.\vec{\omega}\times\vec{v}; \end{cases}$$

(3)

where $\vec{\omega}$ is the angular speed, $\vec{v}$ the translation speed, M the mass of a rigid body and '$\times$' stands for the cross product. The relation (1) leads to 'attractive force', the relation $\vec{F} = M_1.\vec{a}$ (Newton) may leads also to repulsive forces as for the missiles motors and centripedal and Corriolis forces are generally repulsive rotating forces.

To produce locally (on a platform) an inertial force/torque loosing only energy but not masses, Coriolis forces have preferentially to be used.

The mathematical model for the rigid body in movement uses generally moving axis that have their origin on the point around which the rigid body rotate. In 'pure rotation', the linear speed $\vec{v}$ is replaced by the angular speed $\vec{\omega}$ following the relation:

$$\vec{v} = \vec{\omega}\times\vec{r}_1;$$

(4)

where r_1 is the rotation radius. The momentum $m.\vec{v}$ is replaced by the angular momentum $\vec{M}$ defined using the moment of inertia $\vec{I}$ and the angular speed $\vec{\omega}$:

$$\begin{cases} \vec{I} = I_{xx}.\vec{i} + I_{yy}.\vec{j} + I_{zz}.\vec{k} \\ \vec{\omega} = \omega_x.\vec{i} + \omega_y.\vec{j} + \omega_z.\vec{k}; \\ \vec{M} = I_{xx}.\omega_x.\vec{i} + I_{yy}.\omega_y.\vec{j} + I_{zz}.\omega_z.\vec{k}; \end{cases} \quad (5)$$

Here, $\vec{i}, \vec{j}, \vec{k}$ stands for unit vectors. The torques $\vec{L}$ may be obtained as:

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_11, © Springer Science+Business Media B.V. 2010

$$\vec{L} = \frac{d\vec{M}}{dt} = I_{xx}.\frac{d\omega_x}{dt}.\vec{i} + I_{yy}.\frac{d\omega_y}{dt}.\vec{j} + I_{zz}.\frac{d\omega_z}{dt}.\vec{k} + \vec{\omega}\times\vec{M} =$$
$$I_{xx}.\xi_x.\vec{i} + I_{yy}.\xi_y.\vec{j} + I_{zz}.\xi_z.\vec{k} + \vec{\omega}\times\vec{M}; \quad (6)$$

where $\vec{\xi}$ is the angular acceleration. This last relation corresponds to Euler equations of a rigid body with one fixed point:

$$\begin{cases} L_x = I_{xx}.\vec{\xi}_x + (I_{zz} - I_{yy}).\omega_y.\omega_z; \\ L_y = I_{yy}.\vec{\xi}_y + (I_{xx} - I_{zz}).\omega_z.\omega_x; \\ L_z = I_{zz}.\vec{\xi}_z + (I_{yy} - I_{xx}).\omega_x.\omega_y; \end{cases}$$
(6')

It can be remarked that taking into account (3) and (4) it results:

$$\vec{F}_{corr} = 2.M.\vec{\omega}\times\vec{v} = 2.M.\vec{\omega}\times\vec{\Omega}\times\vec{r}_1; \quad (7)$$

which make in evidence the difference between the centripetal $\vec{F}_{cen}$ and Corriolis $\vec{F}_{corr}$ forces: the centripetal force uses a single angular speed but the Corriolis two different angular speeds. Here we have to remark also that a translation may be locally regarded as a rotation with a big enough radius.

II. THE MOTION OF A TOP.

A Top is a simple device that can generate Corriolis forces. When a Top spin fast enough around the axis Pz' where P is its mass center, it can nearly remain into a vertical position in spite of the torque generated by its own weight. The Coriolis forces compensate dynamically this torque.

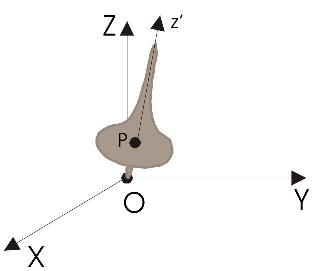


Fig. 1. A Top arrangement.

Following [26], the moving axis in this case are chosen with origin in O, the Oz axis is similar to Pz' as the Top is symmetric about this axis, and the other axis Ox and Oy are let to rotate with an other angular velocity $\vec{\omega}$ than the spin $\vec{s}$ of the Top. Let the angle between OX and Ox be ψ and the angle between OZ and Oz be θ. Then, the angular speeds about the moving axis will be given by:

$$\begin{cases} \omega_x = \frac{d\theta}{dt}; \\ \omega_y =; \frac{d\psi}{dt}.\sin(\theta); \\ \omega_z = \frac{d\psi}{dt}.\cos(\theta); \end{cases}$$
(8)

and by putting $S = s + \omega_z$, the Euler equations becomes:

$$\begin{cases} I_{xx}.\xi_x - I_{xx}.\omega_y.\omega_z + S.I_{zz}.\omega_y = m.g.l.\sin(\theta); \\ I_{xx}.\xi_y + I_{xx}.\omega_x.\omega_z + S.I_{zz}.\omega_x = 0; \\ \frac{dS}{dt}.I_{zz} = 0; \end{cases} \quad (9)$$

where m is the mass of the Top, $PO = l$ and $g \cong 9.81 m/s/s$ the acceleration of the gravity . If S is big enough it result approximately:

$$S.I_{zz}.\omega_y \cong m.g.l = L_x;$$
(10)

which is the well known result for the simple theory of gyroscope.

A top and also a gyroscope have 3 orthogonal axes on which torque and/or angular speeds/acceleration may results or can be applied. As one axis is used for the spin s, only 2 axes are free as is shown in (10). This relation presents also the 'gyrator' property of such a system: an angular speed on Oy axis leads to a torque on the orthogonal axis Ox and evidently, a torque applied on Ox will leads to an angular speed on Oy axis.

Remarks:

- during its movement of precession (and also nutation) with the angular speed ω_y, the Top compensate dynamically the gravitational torque $m.g.l.\sin(\theta)$ due to its own weight. Then, the compensation torque produced by the Top may be also considered as an 'inertial (gravitational) torque'.
- the resulting compensation torque needs a 'strong point' to be realized. This 'strong point' is the point where the peg of the Top lay on the ground. This contact point accepts rotation (theoretically without friction) but must be considered as an 'inertial fixed point in space' as to sustain the static own weight of the Top.
- the nutation movement depend on the initial conditions: initial speed and position in space of Pz' axis. The top of the Top may follow one of the curves shown in the figure 2. The Top (and also gyroscope) motion may be mathematically modeled by Jacobi

elliptical complex functions. Some new properties of these functions was presented in [29]. It can be remarked that these functions models also the field in the air-gap of the electrical motors.

- physically, during the precession movement, the Top fall to the ground due to its own weight. This will accelerate its precession movement. Starting from a given angular speed, the 'gyrator effect' reconverts this angular speed into a torque that tries to compensate the gravitational torque due to the own weight of the Top. The mass of the Top 'filters' these to the precession and nutation movements.

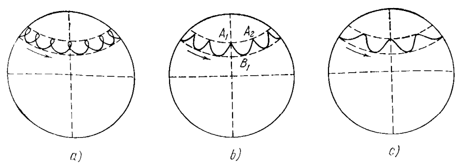


Fig. 2. Nutation and precession possible movement of a Top.

III. PLATFORM EQUIPPED WITH GYROSCOPES.

A gyroscope may be viewed as an improved Top. We will consider a symmetrical gyroscope equipped with 2 gimbals placed so to balance its mass center. Motors are used to sustain the spin of the gyroscope and on the gimbals to apply torques/brakes. The motors on the gimbals may change their speed and direction of the resulting rotation and/or even realize a 'brake'. Torque sensors are placed in the link of the motors with the gimbals to control the applied torques. If the spin is high enough, the gyroscope try to maintain its direction of the Oz axis (on which the spin is applied) versus the stars until torques are not applied on its gimbals.

On a platform in space equipped with such a gyroscope, a torque and/or a forced precession angular speed applied on a gimbals of a gyroscope have to bear on the mass center of this platform if the platform is tacked as a 'charge' for the other gimbals. So, the instantaneous angular speed and/or torque produced by the gyroscope will move from the mass center of the gyroscope to the mass center of the whole platform. During this time, the mass center of the platform will change slightly in space its position (translation) and will begin to rotate. Of course, the direction of Oz axis of the gyroscope will change also versus the stars. If the platform is equipped with 2 such a gyroscopes, one can stop the applied torque on the first gyroscope just before the instantaneous angular speed arrive to the mass center of the platform and apply it to the second gyroscope. This will produce also a (small) translation and the platform will begin to rotate too. A convenient choose of this sequence parameters may lead to a 'snake-like motion' [2], [3], [4]. Of course, as the direction in space of the gyroscopes spin axis change during these operations, one has to restore conveniently the spin axis directions. This may be done by convenient torques applied on a single gimbals of each gyroscope, the other gimbals being free to move. To avoid the necessity of using clutch on the gimbals, we can uses the sensors to insure a 'null torque' delivered by the motors and then, insure an equivalent 'free move' of a desired gimbals.

It can be seen that a convenient sequence of such operations may leads to a translation in space of the platform in any desired direction. A theoretical proof that a translation in space of such a platform may be possible by this means was presented in [5] and [6].

An improvement may be realized by using the 'doublet' [1], [7], [27]. If one gyroscope is placed on the platform as nearest as possible to the other, each gyroscope having the platform as 'charge' for one of its gimbals and the other being used for a common torque (corresponding to the forced precession $\vec{\omega}_y$ for example), it will result a force/torque $\vec{F}$ given by:

$$\begin{aligned}&\vec{F} \cong I_{zz}^{(1)}.\vec{S}_1 \times \vec{\omega}_y \times I_{zz}^{(2)}.\vec{S}_2 = \\ &\vec{\omega}_y.(I_{zz}^{(1)}.\vec{S}_1 \bullet I_{zz}^{(2)}.\vec{S}_2) - I_{zz}^{(2)}.\vec{S}_2.(I_{zz}^{(1)}.\vec{S}_1 \bullet \vec{\omega}_y);\end{aligned} \tag{11}$$

where by '$\bullet$' we have denoted the dot product. In that case, reciprocally, one gyroscope have to bear on the other. Supposing that at the beginning the spins $\vec{S}_1, \vec{S}_2$ are approximately parallel it results:

$$\begin{aligned}&\vec{F} \cong \vec{\omega}_y.(I_{zz}^{(1)}.I_{zz}^{(2)}.S_1.S_2.\cos[angle\ (\vec{S}_1,\vec{S}_2)] - \\ &I_{zz}^{(2)}.I_{zz}^{(1)}.\vec{S}_2.S_1.\omega_y.\cos[angle\ (\vec{S}_1,\vec{\omega}_y)];\end{aligned}$$

(12)

where the term $I_{zz}^{(2)}.I_{zz}^{(1)}.\vec{S}_2.S_1.\omega_y.\cos[angle\ (\vec{S}_1,\vec{\omega}_y)]$ may be neglected for $angle\ (\vec{S}_1,\vec{\omega}_y) \cong \frac{\pi}{2} \pm 2.k.\pi; \to k = 0,1,2,3...$. In that case, the resulting force has the direction of $\vec{\omega}_y$. If the used gyroscopes are (as possible) identical, then:

$$\begin{cases} I_{zz}^{(1)} = I_{zz}^{(2)} = I; \\ S_1 = S_2 = S; \end{cases} \tag{13}$$

and (12) becomes:

$$\begin{aligned}&\vec{F} \cong \vec{\omega}_y.I^2.S^2.\cos[angle\ (\vec{S}_1,\vec{S}_2)] - \\ &\vec{S}_2.I^2.S.\omega_y.\cos[angle\ (\vec{S}_1,\vec{\omega}_y)];\end{aligned} \tag{14}$$

It can be remarked that the maximal value of this force appears 2 times on 360^0 of the revolution angle for Oy axis. To avoid undesired rotations and/or direction change, the torques applied on Oy axis of gyroscopes when brakes (versus the platform) are applied on their Ox axis, have to be applied only for smaller angles than $\pm 15^0$ of the ***angle*** $(\vec{S}_1, \vec{S}_2)$ and ***angle*** $(\vec{S}_1, \vec{\omega}_y)$ around the maximal value of this force. Of course, the equivalent angle corresponding to the angular speed ω_y will depends on the 'charge' applied to the gyroscopes. Lets now consider the case where this equivalent angle correspond exactly to the ***angle*** $(\vec{S}_1, \vec{S}_2)$. Then, the remaining interval of 300^0 for an entire revolution of this equivalent angle corresponding to the angular speed ω_y is to be used to restore conveniently the direction in space of the Oz axis of each gyroscope. This may be done as before by a convenient sequence of torques applied on a single gimbals of each gyroscope, the other gimbals being free to move. Figure 3 shows the relative amplitude variation of the force $\vec{F}$ terms during the angle interval of $\pm 15^0$. As seen, during this time interval, the variations of the force $\vec{F}$ are relatively small, the obtained force will be near to the maximal value of $\vec{F}$.

Like for thermal engines, the obtained force in the case of the gyroscopic doublet will appears only as impulses on a relatively short interval inside a complete revolution period of ω_y. During the remaining time interval dedicated to restore conveniently the direction in space of the Oz axes it may results also relatively small inertial forces/torques. All these forces/torques will be 'filtered' by the entire mass of the platform in the same way in which a 'flying wheel' filters the forces/torques impulses of a thermal engine.

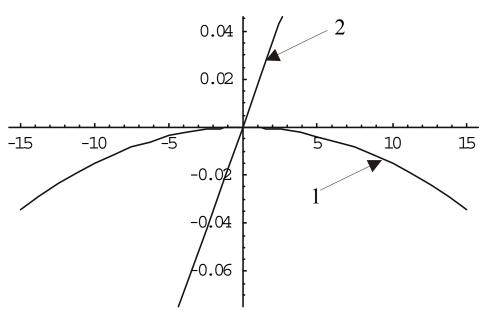


Fig. 3. Relative amplitude variation of 1-> $\vec{\omega}_y . \boldsymbol{I}^2 . \boldsymbol{S}^2 . \cos[\boldsymbol{angle}\ (\vec{S}_1, \vec{S}_2)]$ and 2-> $\vec{S}_2 . \boldsymbol{I}^2 . \boldsymbol{S} . \omega_y . \cos[\boldsymbol{angle}\ (\vec{S}_1, \vec{\omega}_y)$ terms during the angle interval of $\pm 15^0$.

A force applied in opposition to the movement direction of a body will diminish its speed (and kinetic energy too). Then, it may be considered as a 'dynamical brake'. This kind of 'brake' is actually used for the Jet airplanes. If this force is 'inertial', it may replace efficiently the actual brake system for terestrial (car) vehicle when the road is covered by ice, for example. The possibility to change conveniently the direction and value of this inertial force may insure also a good security of cars by diminishing the probabilities of accidents.

It can be observed that two identical gyroscopes placed exactly in 'opposition' on a platform do not produce a gyroscopic effect due to their mutual compensation. This may be used to change easily the direction of the applied force on the platform.

IV. PRACTICAL CONSIDERATIONS.

It has been seen that a platform equipped with 2 identical gyroscopes as described before may generate a force that can translate it. For the case of the 'doublet', the maximum value of this force is given by:

$$\vec{F}_{\max} \cong \vec{\omega}_y . \boldsymbol{I}^2 . \boldsymbol{S}^2; \tag{15}$$

Ideally, if we consider the case where the doublet act for $\pm 15^0$ around the maximal $\vec{F}_{\max}$ and during the remaining interval the delivered forces/torques may be neglected, the mean value of the obtained force on a precession revolution may be given by:

$$F_{mean} \cong \frac{F_{\max} . 60}{360}; \tag{16}$$

Practically, the value of obtained F_{mean} will be less. Of course, the magnitude of this force can be controlled by changing the time interval corresponding to $\pm 15^0$ around the maximal $\vec{F}_{\max}$ and/or changing the angular speed ω_y.

To fix the ideas, we will consider a simple example. Let's considering a system with 2 (ideally identical) gyroscopes, each with a mass of 1 Kg, diameter of 0.2 m, an angular speed for the spin of 20.000 turns/minute and a forced precession angular speed of 200 turns/minute. The obtained mean value force F_{mean} will be of ~ 382.7 Newtons. The internal (or even if desired external) obtained maximal torques using a single gyroscope will be of ~ 219.3 Kg.m. Then, even if the whole platform weights more than 25 Kg, this system will be able to vertically take it off! It results that such a motor system may be able to embark a (relatively) big charge.

For a practical realization of such an inertial motor/brake, the following remarks may be considered:

- from technological point of view, the considered parameters of the gyroscopes are actually feasible and I think that such a system will be mechanically not much more sophisticated than the mechanical system of a helicopter.
- this inertial motor system must include a sensor for the position and direction in space (an inertial platform for example), an electronic processor to insure the convenient sequence of the applied torques/brakes and an energy supply source. For the security of the whole system, a given vehicle must be equipped with at least 2 similar such 'doublets'.

- care must be tacked to balance as much as possible all the internal torques of the system. The remaining torques will be compensated dynamically by the 'doublets' in order to keep the desired direction in space.

- as during the time intervals in which the gyroscopes are exactly in opposition, any change in theirs common direction in space may be done without the display of (big inertial) torques/forces, this fact may be used to change easily the direction in space of delivered force/torque by the 'doublets'.

- to diminish the inherent vibrations dues to the pulsating nature of the resulting force/torque that only the mass of the whole platform will 'filter', a high enough (forced) precession speed have to be used.

- as the gyroscopes are 'gyrators', this system will presents a good efficiency.

- care must be tacked to realize a robust enough 'doublet' system able to withstand its internal forces/torques.

- as the obtained force increase rapidly with the radius of the gyroscopes, a larger size system will display a bigger force even keeping unchanged the other parameters.

- as the gyroscope is a 'gyrator', by an intelligent sequence of applied charge and/or torques/brakes on the gimbals and by using electrical motors/generators, the inertial energy of the ship may eventually be recuperated when a diminishing of its speed is needed. This will insure a more economic trip.

V. CONCLUSIONS

A Top and/or a gyroscope may create inertial torques able to maintain its direction in space but not inertial forces to modify its own weight. As an extension of [2], [3], [4] and in analogy with the Lorenz force [7], a couple of 2 gyroscopes may realize a 'doublet' [1], [7], [27] able to create big enough inertial forces/torques to vertically take off a ship equipped with it. This 'doublet' system has some essential advantages:

- it can be used as motor/brake for terrestrial, submarine and/or even (deep) space ships.

- the mass of the ship which use it will remain nearly the same as it is not necessary to eject masses like for actual Gas-Jets motors and/or rocket engines.

- it can be installed inside the cell of the ship with no external parts as for actual vehicles.

- it may insure a vertical take off directly from the earth ground.

- it may realize accelerations/brakes forces/torques in any desired direction. These accelerations/brakes forces/torques are independent of the ship translation [non-relativist] speed. The brakes may be realized also by recuperating the inertial energy of the ship. This will leads to a more economic trip. All these are based on the gyrator properties of gyroscopes which presents also a good efficiency.

- it can be realized using the actually available technology.

- the energy of the translation obtained through the mean value force $\boldsymbol{F}_{mean}$ results from the rotation energy of the spins ($\vec{\boldsymbol{S}}_1,\vec{\boldsymbol{S}}_2$) and of the torque which realize the forced precession ($\vec{\omega}_y$).

- this system may be also used as active gravitational sensor and/or actuator.

- large ships may presents internal torque/forces that leads to vibrations due to the inertia of the mass far from their motors. This is the case, for example, for the wings of a airplane. One can distribute many synchronized inertial motors into a ship so that this effect may be practically neglected. In this manner the ship may support much higher accelerations.

- to diminish the internal vibrations dues of the pulsate nature of the resulting inertial forces, a high enough (forced) precession speed have to be used.

This inertial motor/brake system may have applications in terrestrial, naval, aviation and even deep space ships. Due to its possibility to obtain higher accelerations, it may also be useful, for example, to diminish the probability of collisions and/or crash by accidents. Many others applications may be envisaged too.

Maybe in the future, 'spin-trap' materials [10] will replace the mechanical gyroscopes like actually the permanent magnets replace the electromagnets in the electrical motors/generators.

REFERENCES.

[1] Robert A. Granger, " Fluid Mechanics", Dover Publications, INC, New York, 1995.

[2] Jean-Francois Longchamp, Dan Ciulin, "Device and method for producing a linear movement and a rotational movement by means of changes in rotational pulses", European patent: IPC: G01C19/00; B63B49/00; (+2), publication info :DE4304129, 18.08 1994.

[3] Jean-Francois Longchamp, Dan Ciulin, "Verfahren und Vorrichtung zur Reduzierung der Nutzbandbreite eines bandbegrenzten Signals durch Kodieren desselben und Verfahren und Vorrichtung zum Dekodieren des bandbegrenzten SignalsVerfahren und Vorrichtung zur Reduzierung der Nutzbandbreite ein... ", European patent: DE58908835D - 1995-02-09.

[4] Jean-Francois Longchamp, Dan Ciulin, "VERFAHREN UND VORRICHTUNG ZUR REDUZIERUNG DER NUTZBANDBREITE EINES BANDBEGRENZTEN SIGNALS DURCH KODIEREN DESSELBEN UND VERFAHREN UND VORRICHTUNG ZUM DEKODIEREN DES BANDBEGRENZTEN SIGNALSVERFAHREN UND VORRICHTUNG ZUR REDUZIERUNG DER NUTZBANDBREITE EIN...", European patent: IPC: H04B1/66; H03C3/06.

[5] Kuang-Yow Lian. Li-Sheng Wang, Li-Chen Fu, "Controllability of Spacecraft Systems in a Central Gravitational Field", IEEE Transaction on Automatic Control, vol. 39, nr.12, December 1994.

[6] Steven K. Blau, "The Force Need Not Be With You: Curvature Begets Motion. ", Physics Today 23, June, 2003.

[7] H. Aspden, "Anti-gravity electronics", Electronics& wireless world, January 1989.

[8] Charles W. Misner, Kip S.Thorne, John Archibald Wheeler, "Gravitation", 1988.

[9] D. Sivoukhine, "Cours de physique générale, Mécanique", Edition MIR Moscou, 1982.

[10] Phil Walker, George Dracoulis, "High-spin nuclear traps", Physics World, February 1994.

[11] A. Ishlinsky, "Orientation, gyroscopes et navigation par inertie", tome I et II, Edition MIR, Moscou, 1984.

[12] Ernest Helmut Brandt, "Theory catch up with flying frog", Physical World, September 1997.

[13] Robert Matthews, Ian Sample, "Breakthrough as scientists beat gravity", The Journal of ideas, issue 4, 02.09. 1996.

[14] Charles P. Poole, Jr., "Electron spin resonance. A comprehensive treatise on experimental techniques", Interscience Publishers, a division of John Wiley&Sons New-York, London, Sidney, 1967.

[15] Jean-Claude Radix, "La navigation par inertie", Presses Universitaires de France,1967.

[16] Feynman and all, "Lectures on physics", Addison-Weley Publishing Company, Inc,1963.

[17] Jean-Claude Radix, "Gyroscopes et gyromètres", Ecole nationale supérieure de l'aéronautique et de l'espace, SUP'AERO', Cepadues editions,1978.

[18] G.Schafer, "The Einstein theory of gravitation and its verifications", Summer course of theoretical physics,EPFL,1990.

[19] *Teldix*, "Ball bearing momentum & reaction wheels+wheel drive electronics", Drallrad+WDE,Heidelberg.

[20] P. Kummel, "Schubprazessor", Stuttgtart 1990.

[21] Canot Albert, "Moteur autopropulseur", European patent application Nr.90470067I.

[22] Earle M. Halvorson, Kurt Schwarz, "Vibration driven vehicle", U.S. patent application, September 29, 1970, Nr. 3.530.617.

[23] Hersey W. Young, "Directional force generator", U.S. patent application, January 19, 1971,Nr.3,555,915.

[24] H. Bouasse, "Cours de mécanique rationnelle et expérimentale", Paris, 1920.

[25] S. A. Agafonov, K. B. Alekseyev, N. V. Nikolayev, "Stability of the control of a plane turn of a spacecraft", J. Appl. Maths Mechs., Vol. 55, No. 1, pp. 136-138, 1991.

[26] Robertt Bruce Lindsay, "Physical mechanics", D. Van Nostrand Company, INC. Princeton, New Jersey, Totonto, London, New York, 1961.

[27] Caius Iacob, " Mathematical introduction in the mechanics of the fluids", edited by the Academy of R.P. Rumania, (in Rumanian), 1951.

[28] D. Ciulin, ”Motor that assures space displacements of a vehicle by means of inertial forces”, Idea for European Project Proposal, ERIMA07’ symposium, Biarritz, 2007 .

[29] D. Ciulin, ”About Sign function and some extensions”, CISSE 2007 Online E-Conference, December 3-12, 2007.

[30] Kip. S. Thorne, ”Black holes and time warp”, W.W. Norton&Company, New York, London 1994.

[31] Roger Penrose, ”The road to reality”, Vintage books, London, 2005.

Active Groups in Complex Mechanisms Structures

Adriana Comanescu, Dinu Comanescu, Ioan Dan Filipoiu, Cristian Gabriel Alionte

University POLITEHNICA
Bucharest, Romania

Abstract*- The paper deals with some active modular groups usually used in the direct structure model of the leg robots or other complex equipment. Such open, un-composed and multi-mobile groups may also be involved in the direct – cinematic or dynamic modeling of the robot displacement for different strategies of motion. There are presented active modular groups with one degree of mobility (the initial active group IAG, the 4R and RTRR active group) and the 5R active modular group with two degrees of mobility. By using the inertia systems and the other external force systems their dynamic models for the determination of all reaction torque components in each (ordinary or active) cinematic pair are elaborated.

Keywords: direct model, active modular group, stepping robot or platform, dynamic control parameter, optimal control design

I. Introduction

The active modular groups may be found in various complex mechanisms with different degrees of mobility [1]. Their involving in the direct model structure of stepping robots is relevant [2]. For example a biped robot or platform can be equipped by two legs each having three degree of mobility (Fig.1). The actuators are placed in the active cinematic pairs namely A, E and F.

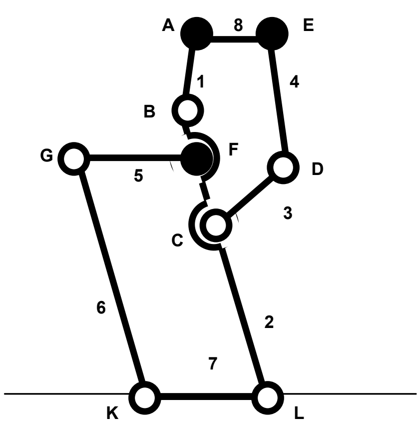


Fig.1. A stepping robot leg solution

During its displacement each leg may alternatively have the output link 7 or 8 attached to the platform. When the output link is 7, the basis of the mechanism is the 8 link, so that the leg system has the direct structure model presented in Fig.2.

In the case in which the output link is 8 the basis of the mechanism is the link 7 rigidly attached to the soil. The direct structural model corresponding to this phase of the platform displacement is presented in Fig.3.

In Fig.2 and Fig.3 there are mentioned the active modular groups (AMG) with one or two degrees of mobility and the passive modular ones (PMG) with zero degree of mobility.

The reason of this concrete example is that of illustrating the use of such active modular groups in the structure of different mechanical systems.

In function of their degrees of mobility these modular groups have active rotation or prismatic pairs assimilated to the actuators.

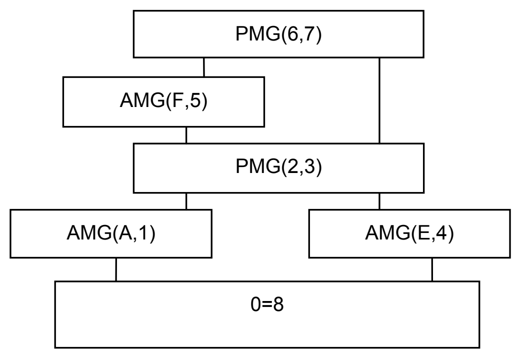


Fig. 2. The direct structure model for the 7 output link

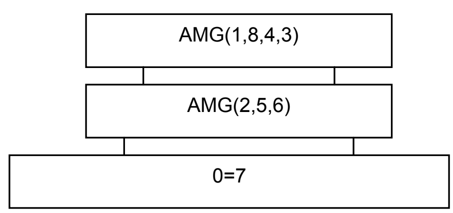


Fig. 3. The direct structure model for the 8 output link

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_12, © Springer Science+Business Media B.V. 2010

In the next sections of the paper there are presented the mathematical models for some modular groups in order to determine the components of the reaction torque of each pairs. The following notations are used:

- $T_i(xT_i, yT_i)$ - the link mass centre,
- $\tau_i(X_i, Y_i, CM_i)$ – the equivalent torque of the inertia and exterior force systems applied in T_i for the i link.

The reaction torque components are established by using the dynamic equilibrium equations.

II. THE DYNAMIC MODULES FOR THE ACTIVE GROUPS WITH ONE DEGREE OF MOBILITY

In this section there are shortly presented three active modular groups, which are the initial active modular group (IAMG – Fig. 4), 4R group (Fig.5 and Fig. 6) and RTRR group (Fig.7 and Fig. 8).

The initial active modular group model (Fig.4) has the input parameters mentioned in TABLE I. From the equilibrium dynamic equations there are obtained the output parameters mentioned in TABLE II.

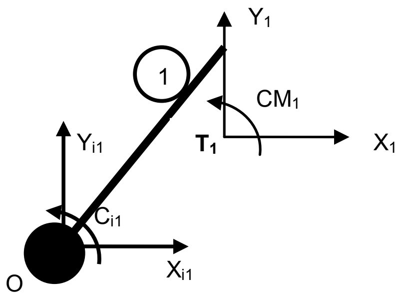


Fig.4. The initial active modular group model

TABLE I
THE INPUT PARAMETERS FOR THE IAMG

The coordinates of the O active pair	$O(XO,YO)$
The coordinates of the T $_i$ mass centre of the 1 link	$T_1(xT_1, yT_1)$
The equivalent torque of the inertia and external force systems applied to the 1 link	$\tau_1(X_1, Y_1, CM_1)$

TABLE II
THE OUTPUT PARAMETERS OF THE IAMG

The reaction torque of the O active pair	$X_{i1} = -X_1$ $Y_{i1} = -Y_1$ $C_{i1} = -[CM_1 - (yT_1 - yO)X_1 + (xT_1 - xO)Y_1]$

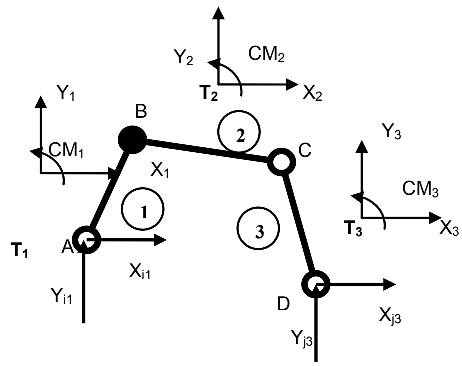


Fig.5. The 4R active modular group

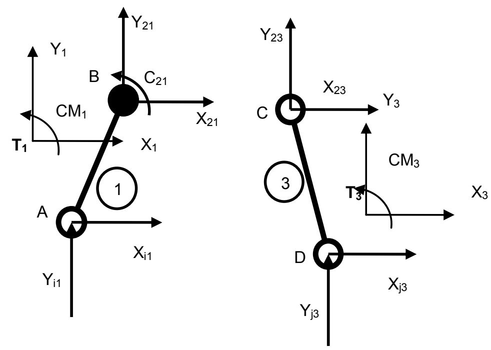


Fig.6. The isolated links of 4R active group

TABLE III
THE INPUT PARAMETERS FOR THE 4R ACTIVE GROUP

The cinematic pair coordinates	$A(XA,YA)$ $B(XB,YB)$ $C(XC,YC)$ $D(XD,YD)$
The mass centre coordinates	$T_1(XT_1, YT_1)$ $T_2(XT_2, YT_2)$ $T_3(XT_3, YT_3)$
The equivalent torque of the inertia and external force systems for each cinematic link	$\tau_1(X_1, Y_1, CM_1)$ $\tau_2(X_2, Y_2, CM_2)$ $\tau_3(X_3, Y_3, CM_3)$

The 4R modular group is given in Fig.5 and its input parameters in TABLE III are deduced using isolated links of 4R, as shown Fig. 6.

As to the active group with three mobile links there are two possibilities that are the active pair being a rotation or prismatic pair. Consequently, there are two such modular groups.

TABLE IV
THE OUTPUT PARAMETERS OF THE 4R ACTIVE GROUP

The reaction torque of the D pair	$A\cdot\begin{bmatrix}X_{j3}\\ Y_{j3}\end{bmatrix}=B$ $A=\begin{bmatrix}-(yD-yC) & xD-xC\\ -(yD-yA) & xD-xA\end{bmatrix}$ $B=\begin{bmatrix}-[CM_3-(yT_3-yC)X_3+(xT_3-xC)Y_3]\\ -[CM_3-(yT_3-yA)X_3+(xT_3-xA)Y_3+\\ CM_2-(yT_2-yA)X_2+(xT_2-xA)Y_2+\\ CM_1-(yT_1-yC)X_1+(xT_1-xC)Y_1]\end{bmatrix}$
The reaction torque of the A pair	$X_{i1}=-(X_{j3}+X_1+X_2+X_3)$ $Y_{i1}=-(Y_{j3}+Y_1+Y_2+Y_3)$
The reaction torque of the C pair	$X_{23}=-(X_{j3}+X_3)$ $Y_{23}=-(Y_{j3}+Y_3)$
The reaction torque of the B pair	$X_{21}=-(X_{i1}+X_1)$ $Y_{21}=-(Y_{i1}+Y_1)$ $C_{21}=-[-(yA-yB)X_{i1}+(xA-xB)Y_{i1}$ $+CM_1-(yT_1-yB)X_1+(xT_1-xB)Y_1]$

Each ordinary rotation pair has for its reaction torque only two force components. An active rotation pair as B in Fig.6 has a reaction torque with three components, two forces as the previous ones and a moment, a interaction of two fields (of various nature) existing between the adjacent links. One may obtain these component torques by applying the relations from TABLE IV.

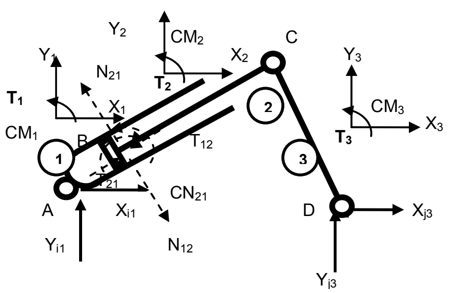


Fig.7. The RTRR active modular group

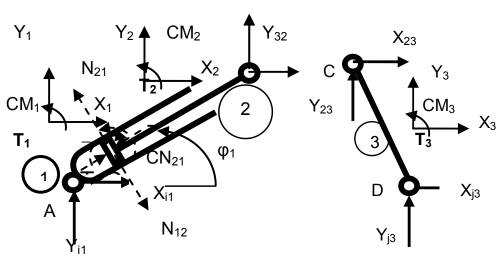


Fig.8. The isolated links of RTRR active modular group

The RTRR active group is shown in Fig.7, its input parameters being the same as those from TABLE III are deduced using isolated links of RTRR, as shown Fig. 8.

The torque components from the D pair are given from the system (1).

$$A\begin{bmatrix}X_{j3}\\ Y_{j3}\end{bmatrix}=B \qquad (1)$$

$$A=\begin{bmatrix}-(yD-yC) & xD-xC\\ -(yD-yA) & xD-xA\end{bmatrix}$$

where

$$B=\begin{bmatrix}-[CM_3-(yT_3-yC)X_3+(xT_3-xC)Y_3]\\ -[CM_3-(yT_3-yA)X_3+(xT_3-xA)Y_3+\\ +CM_2-(yT_2-yA)X_2+(xT_2-xA)Y_2+\\ +CM_1-(yT_1-yA)X_1+(xT_1-xA)Y_1]\end{bmatrix} \qquad (2)$$

The components for the A external pair are obtained from (4), so that there are:

$$\begin{aligned}X_{i1}&=-(X_{j3}+X_1+X_2+X_3)\\ Y_{i1}&=-(Y_{j3}+Y_1+Y_2+Y_3)\end{aligned} \qquad (3)$$

The components for the C internal rotation pair are:

$$\begin{aligned}X_{23}&=-(X_{j3}+X_3)\\ Y_{23}&=-(Y_{j3}+Y_3)\end{aligned} \qquad (4)$$

and those for the B active pair are expressed by (5).

$$\begin{aligned}T_{21}&=-[(X_{i1}+X_1)\cos(\varphi_1)+(Y_{i1}+Y_1)\sin(\varphi_1)]\\ N_{21}&=-[-(X_{i1}+X_1)\sin(\varphi_1)+(Y_{i1}+Y_1)\cos(\varphi_1)]\\ CN_{21}&=-[-(yA-yB)X_{i1}+(xA-xB)Y_{i1}+\\ &+CM_1-(yT_1-yB)X_1+(xT_1-xB)Y_1]\end{aligned} \qquad (5)$$

which allow establishing their variation relative to time.

III. THE DYNAMIC MODULE FOR THE 5R ACTIVE GROUP WITH TWO DEGREES OF MOBILITY

The 5R active group may be found in the direct model of the leg when the fixed link is 7. Its structure is presented in Fig.9 and the input parameters are in TABLE IV using isolated links, presented in Fig. 10.

TABLE IV
THE INPUT PARAMETERS OF THE 5R ACTIVE GROUP

The cinematic pair coordinates	$A(XA,YA)$ $B(XB,YB)$ $C(XC,YC)$ $D(XD,YD)$ $E(XE,YE)$
The mass centre coordinates	$Ti(xT_i,yT_i)$ where $i=1\div4$
The equivalent torque of the inertia and external force systems for each cinematic link	$\tau_i(X_i,Y_i,CM_i)$

The determination of the dynamic characteristics for the active pairs C and D makes possible to design and to achieve the active control on the prismatic actuators placed in the construction of each leg of the platform. Such structural active groups for the legs may be identified in some other constructive platform solutions.

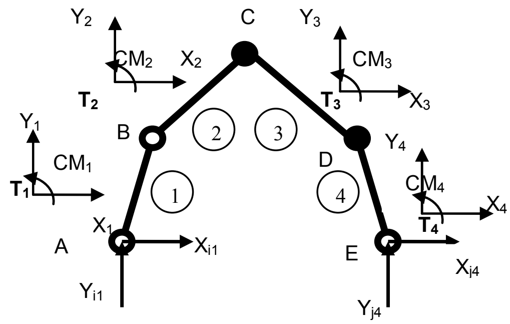


Fig.9. The 5R active modular group

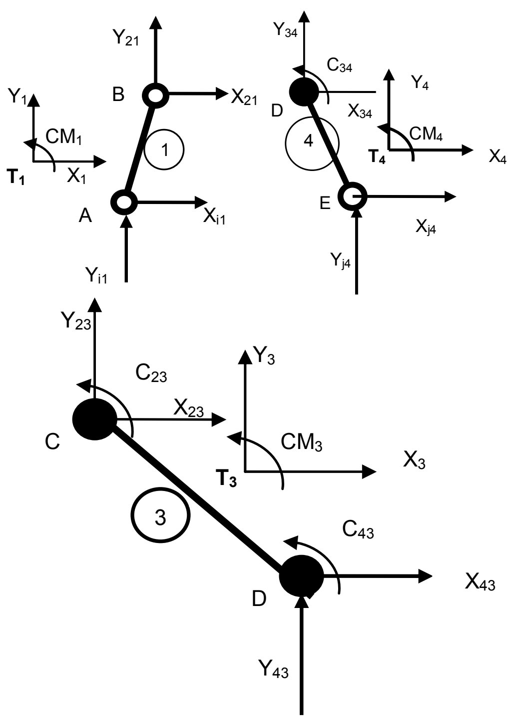


Fig.10. The isolated links of the 5R active modular group

The reaction torque in the A pair may be deduced from the following system

$$A \cdot \begin{bmatrix} X_{i1} \\ Y_{i1} \end{bmatrix} = B \tag{6}$$

where

$$A = \begin{bmatrix} -(yA - yB) & xA - xB \\ -(yA - yE) & xA - xE \end{bmatrix} \tag{7}$$

$$B = \begin{bmatrix} -[CM_1 - (yT_1 - yB)X_1 + (xT_1 - xB)Y_1] \\ -[CM_4 - (yT_4 - yE)X_4 + (xT_4 - xE)Y_4 \\ + CM_3 - (yT_3 - yE)X_3 + (xT_3 - xE)Y_3 + \\ + CM_2 - (yT_2 - yE)X_2 + (xT_2 - xE)Y_2 + \\ CM_1 - (yT_1 - yE)X_1 + (xT_1 - xE)Y_1 \end{bmatrix}$$

The reaction torque for the E external pair is given by

$$\begin{aligned} X_{j4} &= -(X_{i1} + X_1 + X_2 + X_3 + X_4) \\ Y_{j4} &= -(Y_{i1} + Y_1 + Y_2 + Y_3 + Y_4) \end{aligned} \tag{8}$$

and the relations for the B internal pair torque are

$$\begin{aligned} X_{21} &= -(X_{i1} + X_1) \\ Y_{21} &= -(Y_{i1} + Y_1)] \end{aligned} \tag{9}$$

The dynamic parameters for the D active pair torque may be determined as follows:

$$\begin{aligned} X_{34} &= -(X_{j4} + X_4) \\ Y_{34} &= -(Y_{j4} + Y_4) \\ C_{34} &= -[-(yE - yD)X_{j4} + (xE - xD)Y_{j4} \\ &+ CM_4 - (yT_4 - yD)X_4 + (xT_4 - xD)Y_4] \end{aligned} \tag{10}$$

IV. CONCLUSION

The active modular group concept is useful for the design of the systems with multi-degrees of mobility. Its use has a major importance for the determination of the system critical solutions corresponding to their direct (cinematic, dynamic) model.

REFERENCES

[1] Comanescu, D., Comanescu, A., "Displacement Models for Star-Stop Platforms", *Mecatronica*, ISSN 1583-7653, p.27-31,2004;

[2] Comanescu, A., Comanescu, D., "Bi-Mobile Structures Optimally Associated to Inverse Modules from Robotics and Mechatronics", *Rev. Constr.de Masini*, no.6, pp.22-26,1998.

Dimensions of Online Behavior: Implications for Engineering E-Learning

Genevieve Marie Johnson
Department of Psychology, Grant MacEwan College Canada

Julia Ann Johnson
Department of Mathematics and Computer Science, Laurentian University, Canada

Abstract **E-students vary in terms of *online behavior*, defined as organized (e.g., web search) and disorganized (e.g., browse) interactions with both human (e.g., chat) and nonhuman (e.g., database) elements in online environments. Online behavior is conceptualized in terms of the dimensions of *sociability* (human connection motives), *utility* (efficiency orientation), and *reciprocity* (the need for cognitive stimulation and active involvement when using the Internet). Individuals from various disciplines completed the *Brief Test of Online Behavior* and received scores on the three dimensions of online behavior. Although variation was apparent, the professional engineer was not distinguished by his scores on sociability, utility, or reciprocity, suggesting that all three dimensions of online behavior should be considered in e-learning.**

Index Terms **- Online behavior, Internet use, e-learning, sociability, utility, reciprocity, engineering education**

I. Introduction

Students differ on a range of cognitive (intellectual) and psychosocial (personal) characteristics. Research suggests that mismatch between instructional presentation and individual characteristics may lead to lack of student motivation, and may cause attrition [1], [2]. Holt and Solomon [3] noted that because engineering education relies heavily on problem solving, it tends to exclude creative and highly social students from effective learning. It also limits opportunities of all learners to develop the skills required for proficiency in two other key areas of engineering: design and invention (requiring creativity) and business management (requiring social skills). There is "considerable evidence from previous studies to show that faculty assumptions about typical … engineering students are inaccurate and that a wide range of learning styles will actually exist within any engineering class [4]." Increasingly, conventional (face-to-face) engineering education has focused on accommodating learner differences by using a variety of instructional strategies and techniques [5], [6]. However, as computer technology revolutionizes instruction, accommodating student differences in e-learning environments is of increasing concern.

II. E-Learning, E-Teaching, and E-Students

E-learning is a general term used to describe education where the medium of instruction is computer technology [7]. In practice, e-learning is synonymous with Internet-mediated instruction [8]. E-teaching processes typically include a variety of applications of Internet tools, for example, online study groups [9], web-released tests and quizzes [10], and instructional use of the communication tools [11]. In companies, e-learning refers to strategies that use the company network to deliver training courses to employees. In most universities, e-learning refers to a mode of instruction where students rarely, if ever, attend face-to-face on-campus courses, because they study online.

E-learning is aggressively promoted in engineering as a mechanism of dealing with projected and current work force shortages [12]. In some cases, e-teaching is more effective than face-to-face instruction [7], [13]. E-learning is well-suited to technically oriented, project management skills rather than less tangible interpersonal skills [14]. However, mismatch between e-teacher presentation of Internet learning tools and individual student characteristics may lead to lack of e-student motivation and interest, and may cause attrition in e-learning courses and programs [15], [16]. E-student characteristics are an important consideration in e-learning pedagogy.

A. E-Student Characteristics

Research has demonstrated that a range of student cognitive (intellectual) and psychosocial (personal) characteristics influence e-learning outcomes. Johnson [17] found that increased student WebCT use in a hybrid learning environment was associated with inadequate peer relations. Cadieux [18] reported that college students in face-to-face learning groups had stronger feelings of trust and more effective interactions than students in online learning groups, although "no significant relationship was found between sense of community and course grades." McConnell [19], examining the dynamics of collaborative e-learning groups, noted that students' sense of personal identity, control, and security were related to effective group functioning. Johnson, Howell, and Code [13] proposed a number of student characteristics (e.g., anxiety and self-efficacy) by which e-learning may be both positively and negatively affected. Reportedly, aspects of classroom climate (e.g., perception of teacher and peer support, academic alienation, orientation to learning) influence student behavior in online learning environments [20].

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_13, © Springer Science+Business Media B.V. 2010

The personality dimension of introversion-extroversive has also been linked to e-student satisfaction and success in online learning environments [21], [22]. Extroverts tend to focus on the outer world of people and external events while introverts "focus on their own inner world of ideas and experiences [23]." Soles and Moller [24] proposed that synchronous Internet communication is best suited to extroverts while asynchronous Internet communication is best suited to introverts. Daughenbaugh, Daughenbaugh, Surry, and Islam [25] compared student satisfaction in online and traditional computer science courses in terms of personality characteristics. "The data indicated that students rated as extroverts, rather than introverts, showed a stronger preference for the ways in which information is presented in online courses [25]."

Learning style refers to "characteristic cognitive, affective, and psychological behaviours that serve as relatively stable indicators of how learners perceive, interact with, and respond to the learning environment [26]." Numerous studies report that student e-learning outcomes are influenced by learning style [27]-[29]. Bajraktarevic, Hall, and Fullick [30] found that students demonstrated better task performance when online activities were matched with their learning styles. Mehlenbacher, Miller, Covington, and Larsen [31] concluded that reflective learners (those who preferred personal introspection and working alone) were more successful in web-based courses than active learners (those who preferred to engagement in physical activity and group discussion). "During the past ten years, a great number of studies have found that users' cognitive styles significantly influence their reaction to the user interface in terms of user control, multiple tools, and nonlinear interaction [32]." Based on empirical data, Graff [15] recommended designing systems that are less segmented for analytic individuals and more segmented for creative individuals.

Thus, it appears that a range of student cognitive and psychosocial variables are implicated in e-learning outcomes (i.e., student achievement, satisfaction, and persistence). Further, it has been argued that online environments may be less accommodating to learner differences than traditional face-to-face classrooms [33], [34]. Effective instructional practice in traditional classrooms may not necessarily generalize to e-learning environments. Indeed, the human behavior that occurs in online environments is increasingly recognized as complex and idiosyncratic [35]. Accommodation of e-learner cognitive and psychosocial differences requires precise description of the phenomena of interest, that is, *online behavior*.

III. The Salient Features of Online Behavior

Online behavior refers to organized (e.g., web search) and disorganized (e.g., browse) interactions with both human (e.g., in chat rooms) and nonhuman (e.g., database) elements in online environments [35]. Johnson and Kupla [35] proposed a user typology based on the salient features or distinguishing attributes of online behavior. Synthesis of the literature supports the conclusion that online behavior is reasonably conceptualized by the dimensions of *sociality*, *utility,* and *reciprocity.*

A. The Sociability Dimension of Online Behavior

Joinson [36] maintained that online behavior reflects a variety of psychological motives including affiliation (e.g., social connectedness). Herring [37] suggested that sociability is an essential feature of computer mediated communication. Reportedly, 57% of all Internet use relates to communication (i.e., email, instant messaging, and chat) [38]. The *Web Motivation Inventory* [39] identifies communication as a primary motive for using the Internet. Variation in online social interaction has been used to explain individual differences in patterns of Internet use [40], [41].

Online behavior, however, can also be asocial [42], [43]. Users engage in solitary online activities such as playing games and browsing; Internet use may lessens face-to-face social exchange [44]. Waldvogel [45] noted that, in the workplace, some individuals use email to avoid face-to-face contact. Antisocial uses of the Internet use such as flaming [46] and bullying [36] are not uncommon.

As illustrated in Figure 1, the sociability dimension of online behavior ranges from efforts to maximize to efforts to minimize social exchange. To illustrate, a social phobic demonstrates extremely low online sociability by using the Internet to minimize or avoid social contact (e.g., online shopping). Moderate sociability occurs when Internet applications are used to maintain relationships (e.g., email); high levels are evidenced by individuals who use the Internet to establish new relationships (e.g., online dating).

Fig. 1. The sociability dimension of online behavior.

B. The Utility Dimension of Online Behavior

The Internet is frequently conceptualized as a tool that facilitates the execution of complex human intention (e.g., storing, organizing, and retrieving information) [47], [48]. Papacharissi and Rubin [49] concluded that many users have an instrumental orientation to the Internet, "characterized by utility, intention, selectivity, and involvement." In online environments, the instrumental orientation stresses drive for efficiency and effectiveness [50]. Volk and Kraft [51] noted that individuals vary in their perception of the Internet as a mechanism of utility versus a source of recreation. Johnson [52] reported that 77.8% of college students consider the Internet a convenience (utility) while 17.8% consider it a source of amusement.

Online utility is conceptualized as falling on a continuum that ranges from highly efficient execution of specific tasks (e.g., renew library books) to behavior such as surfing and browsing which, according to Kalbach [53], is largely unfocused and unproductive. Individual differences in online utility reflect individual differences in user characteristics [54]. Figure 2 presents the utility continuum of online behavior. The higher the instrumental orientation of the user, the more precise and directed the online behavior.

Fig. 2. The utility dimension of online behavior.

C. The Reciprocity Dimension of Online Behavior

In addition to differences in sociability and utility, online behavior varies as a function of reciprocity (i.e., reciprocal action and reaction). Different sites support [55], and different users require [56], variation in sensory stimulation and active involvement. To illustrate, an individual may use the Internet to read text (low reciprocity), transfer funds (moderate reciprocity), or chat with friends (high reciprocity). Internet applications associated with high levels of reciprocity involve high levels of human interaction and/or interaction with complex interfaces [57]. Volk and Kraft [51] concluded that users vary in extent of website involvement and that high levels of e-commerce reciprocity were associated with increased probability of the user making a purchase.

Figure 3 presents the reciprocity dimension of online behavior. Reciprocity in online environments reflects behavior that ranges from user preference for simple interfaces, with only one browser window open, in which limited text and simple images are interpreted to, for example, multi-modal multi-player games in which multiple identities are executed in virtual reality [58]. Individual differences in reciprocity in online environments reflect individual differences in user cognitive and psychosocial variables [59]. For example, Johnson [16] reported that active and reflective learners have predictable preferences for web-based course tools. As is the case in conventional classrooms, active learners prefer high levels of stimulation and involvement; reflective learners prefer minimal distraction and limited social involvement [60].

Fig. 3. The reciprocity dimension of online behavior.

D. Validating the Dimensions of Online Behavior

Principal component and factor analyses are statistical techniques applied to a set of variables to discover which subsets of variables are coherent and relatively independent of one another. Variables that correlate with one another and that are also largely independent of other subsets of variables combine to form factors or principal components. Factors are assumed representative of the underlying processes responsible for the correlations among variables.

Johnson [52] conducted a study in which college students completed a rating scale that assessed patterns of Internet use. Five items assessed general Internet use (e.g., I use the Internet to help with school work), five items assessed Internet communication (e.g., I use email), five items assessed website access (e.g., I use search engines), and four items assessed online game playing (e.g., I play strategy games online). Students rated each item on a 5-point scale ranging from 1 (never) to 5 (every day or almost every day). Results describe college students, with rare exception, as Internet users.

Principal component analysis of rating scale items revealed three distinct patterns of online behavior in college students [52]. The first principal component loaded positively on all Internet use rating scale items and heavily on most items. The two weakest coefficients were for student rating of the items *I online date* and *I play sports games online*. Such a pattern of coefficient strength was interpreted as describing *integrated-Internet-use* (i.e., extensive use of most applications). The second principal component loaded positively and heavily on the online gaming items but negatively or marginally on all other items. Such a combination of coefficients was interpreted as reflecting a pattern of online behavior described as *games-only use* (i.e., limited use of most applications and extensive online gaming). The third principal component loaded positively and heavily on two items (*I visit chat rooms* and *I online date*) but negligibly on most other uses of the Internet. Such a pattern of coefficients was interpreted as reflecting a pattern of online behavior described as *dating-only use*, that is, limited use of the Internet except for online dating (dating sites typically include chat feature) [61].

The online behavior patterns of integrated-Internet use, games-only use, and dating-only use [52] are easily situated on the continuums of online sociability, utility, and reciprocity. In general, dating-only use is high on sociability and reciprocity but low on utility. Games-only use is low on utility, moderate to high on reciprocity, and moderate to low on sociability. Integrated-Internet-use likely reflects behavior that is balanced in terms of sociability, utility, and reciprocity.

LaRose and Eastin [62] employed structural equation modeling techniques to determine the relationship between Internet use and gratification. Adult Internet users completed a questionnaire in which 29 commonly reported gratifications of Internet use (e.g., feel entertained) were phrased as outcome expectations (i.e., when using the Internet, how likely are you to __). Users rated each statement on a scale from that ranged from 1 (very unlikely to occur) to 7 (very likely to occur). Subject ratings of outcome expectations factored into six gratification or incentive categories: *activity* (e.g., hear music I like), *monetary* (e.g., find bargains on products and services), *novel* (e.g., obtain information that I can't find elsewhere), *social* (e.g., maintain a relationship I value), *self-reactive* (e.g., relieve boredom), and *status* (e.g., find others who respect my views).

These six Internet user gratification factors [62] are easily situated on the dimensions of sociability, utility, and reciprocity. In general, expected monetary outcomes are high on utility and reciprocity but low on sociability. Expectation of novel outcomes is high on utility (e.g., get immediate knowledge of big news events) and reciprocity (e.g., find new interactive features) but moderate to low on sociality. Social outcomes are high on sociability (e.g., get support from others) and, since communication applications are required for social exchange, high to moderate on reciprocity and utility. Statements that factored into self-reactive outcomes reflect undirected online behavior (e.g., a way to pass time) and as such are low on utility but vary as a function of sociability and reciprocity. Status outcome statements are often vague (e.g., improve my future prospects in life) implying moderate utility and, since communication applications are involved (e.g., provide help to others), moderate to high on sociability and reciprocity.

Online behavior is conceptualized by the orthogonal dimensions of *sociability* (human connection motives), *utility* (efficiency orientation), and *reciprocity* (the need for cognitive stimulation and active involvement when using the Internet) [35]. As presented in Figure 4, online sociability, utility, and reciprocity are presented as continuums that intersect to create a three-dimensional space appropriate to theoretical and applied investigation of online behavior. A characterization of online behavior emerges as a triple <s, u, r> where *s* lies somewhere on the sociability dimension and *u* and *r* somewhere on the utility and reciprocity dimensions, respectively.

Fig. 4. Space created by the three dimensions of online behavior.

E. *The Brief Test of Online Behavior*

The *Brief Test of Online Behavior* (BTOB), appropriate for adult Internet users, measures sociability, utility, and reciprocity [35]. As summarized in Table 1, the BTOB includes 15 items; 5 items assess individual position on the three dimensions of online behavior. Users rate items on a 5-point scale ranging from 0 (never) to 4 (always). BTOB items are based on review of the literature and have high face validity.

To reduce biased responding, some items of the BTOB are reverse scored [35]. Specifically, prior to determination of user score on each of the three dimensions of online behavior, ratings for items 1, 4, 6, 8, 10, and 13 (as presented in Table 1) are reversed (i.e., never = always, rarely = often, and so on). Once these six items are reversed scored, dimension scores are determined by summation of scores on the five dimension rating scale items. User sociability, utility, and reciprocity scores potentially range from 0 to 20 (5 items, each with a score from 0 to 4). The greater the sociability, utility, and reciprocity scores, the greater the extent to which the individual is characterized by those dimensions of online behavior. The BTOB permits exploration of individual variation on the three dimensions of online behavior. Such variation has implications for e-learning. For example, does online behavior vary as a function of discipline?

TABLE I
ITEMS IN THE *BRIEF TEST OF ONLINE BEHAVIOR*

Sociability	Utility	Reciprocity
1. I email people I hope to avoid.	6. I spend hours online without accomplishing anything.	11. I ___ have several windows open on my PC at the same time.
2. I ____ add new contacts to my online accounts.	7. The Internet helps me get the job done.	12. Animation ___ improves web pages.
3. I meet new people online.	8. I use the Internet to have fun.	13. I ___feel that websites provide too much information.
4. Using the Internet interferes with my social life.	9. The Internet saves me time.	14. When I use the Internet, my hand is ___ on an input device
5. The Internet helps me keep in touch with friends or family	10. Time seems to fly when I browse and surf).	15. I __ instant message or chat more than I email.

IV. Disciplinary Variation in Online Behavior

Eight individuals from various professions or disciplines completed the BTOB (a sample of convenience). Four of the professionals were from technical fields (i.e., physics, mathematics, computer science, and engineering) and four were from people-oriented professions (i.e., social work, psychology, law, and sociology). Four of the professionals were female and four were male. Three of the professionals described their career stage as *early*, four as *mid*, and one as *late*. Table 2 presents the sociability, utility, and reciprocity scores for each of these eight professionals.

TABLE 2
ONLINE BEHAVIOR SCORES FOR VARIOUS PROFESSIONALS

Professional	Sociability	Utility	Reciprocity
Social Worker	13	11	13
Lawyer	12	16	11
Psychologist	10	17	10
Sociology	9	10	5
Computer Scientist	6	12	15
Mathematical	13	14	5
Physicist	12	12	10
Engineer	15	15	13

Scores potentially range from 0 – 20; the greater the score, the further right the individual falls on dimension of online behavior.

Individual scores on the online behavioral dimensions of sociability, utility, or reciprocity (Table 2) were compared with group descriptive statistics (Table 3). Six online behavior scores, distributed across five disciplines, were more than 1 standard deviation from the mean. That is, on the online behavioral dimension of *sociability* (human connection motives), the computer scientist scored low and the engineer scored high. Both the lawyer and the psychologist scored high on the online behavioral dimension of utility (efficiency orientation), and the sociologist scored low and the computer scientist scored high on *reciprocity* (the need for cognitive stimulation and active involvement when using the Internet). Thus, while differences in patterns of individual online behavior were apparent, such variation was distributed across disciplines, failing to support stereotypical assumptions about the cognitive and psychosocial characteristics of engineers (e.g., low need for social interaction and high need for challenge) [63]. In fact, there were no significant differences in mean sociability, utility, and reciprocity scores for professionals grouped according to technical field (i.e., physics, mathematics, computer science, and engineering) or people-oriented professions (i.e., social work, psychology, law, and sociology).

TABLE 3
DESCRIPTIVE STATISTICS FOR ONLINE BEHAVIOR SCORES

Dimension	Range	Mean	SD
Sociability	9.0	11.25	2.82
Utility	6.0	13.63	2.20
Reciprocity	10.0	10.25	3.84

V. IMPLICATIONS FOR ENGINEERING E-LEARNING

Accommodating student differences in online learning environments requires instructional consideration of student variation in patterns of online behavior. Such consideration results in online instruction that matches e-students position in the space created by the intersecting dimensions of sociability, utility, and reciprocity. For example, presentation of information and student skill development should include highly social exercises (e.g., online group problem solving) as well as more solitary learning events (e.g., web-released individual problem solving). E-learning activities should also allow for student variation in online behavior characterized by a utility orientation. For example, Internet assignment submission and grades access may be important to students who are high on utility; informal online study groups and instructional games may be important to students who are low on utility. Finally, e-learners demonstrate differences in online reciprocity or the need for cognitive stimulation and active involvement when using the Internet. E-learning activities might include, for example, both asynchronous discussion (for students low on reciprocity) and synchronous discussion (for students high on reciprocity).

VI. LIMITATIONS AND FUTURE RESEARCH

The BTOB has not been validated (i.e., information on reliability and validity are not available) [35]. Subsequent research may confirm that the items presented in the BTOB accurately measure the three proposed dimensions of online behavior. Additionally, the sample of professionals was extremely limited. It is possible, perhaps likely, that each professional was not representative of his/her discipline. Indeed, this paper has argued that, in the case of engineering education, student stereotypes are currently unsubstantiated. Future research, with large random samples, may not confirm the current findings. The three dimensions of online behavior require application in instructional practice. Are e-learning outcomes improved when e-teachers plan for student variation in online sociability, utility, and reciprocity?

REFERENCES

[1] R. Felder and L. Silverman. (1988). Learning and teaching styles in engineering education. *Journal of Engineering Education* [Online]. *78(7)*. pp. 674-681. Available: http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/LS-1988.pdf

[2] R. Felder. (1996). Matters of style. *Prism (American Society of Engineering Education)* [Online]. *6(4)*. pp. 18-23. Available: http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/LS-Prism.htm

[3] J. Holt and F. Solomon. (1996). Engineering education - The way ahead. *Australasian Journal of Engineering Education* [Online]. *7(1)*, 47-60.

[4] J. Mills, M. Ayre, D. Hands, and P. Carden, "Learning about learning styles: Can this improve engineering education?" *Mountain Rise*, vol. 2, Fall 2005.

[5] R. Felder, G. Felder, and E. Dietz. (2002). The effects of personality type on engineering student performance and attitudes. *Journal of Engineering Education* [Online]. 9(1). pp. 3-17. Available: http://www4.ncsu.edu/unity/lockers/users/f/felder/public/ILSdir/ILS_Validation(IJEE).pdf

[6] R. Felder and J. Spurlin. (2005). Applications, Reliability, and Validity of the Index of Learning Styles. *Int. J. of Engineering Education* [Online]. *21(1)*. pp. 103-112. Available: http://www.ncsu.edu/felder-public/ILSdir/ILS_Validation(IJEE).pdf

[7] G. M. Johnson and J. A. Johnson, "Computer technology and human learning: Review of recent quantitative syntheses," *Technology, Instruction, Cognition, and Learning*, vol. 4, pp. 287-301, 2006.

[8] G. M. Johnson, "Restricted versus unrestricted learning: Synthesis of recent meta-analyses," *AACE J.,* vol. 15, pp. 267-278, 2007.

[9] G. M. Johnson, "Online study groups: Reciprocal peer questioning vs. mnemonic devices," *J. of Educational Computing Res.,* vol. 35, pp. 83-96, 2006.

[10] G. M. Johnson, "Optional online quizzes: College student use and relationship to achievement," *Canadian J. of Learning and Technology,* vol. 32, pp. 105-118, 2006.

[11] G. M. Johnson, "Synchronous and asynchronous text-based CMC in educational contexts: A review of recent research," *TechTrends,* vol. 50, pp. 46-53, 2006.

[12] K. Nigim, G. T. Heydt, and J. Palais, "E-Learning Opportunities for Electric Power Engineers," *IEEE Trans. on Power Syst.,* vol. 22, no. 3, pp. 1382-3, Aug. 2007.

[13] G. M. Johnson, A. J. Howell, and J. R. Code, "Online discussion and college student learning: Toward a model of influence," *Technology, Pedagogy and Education,* vol. 14, pp. 61-75, 2005.

[14] R. Ellis, T. Thorpe, and G. Wood, "E-learning for project management," *Proc. of the Institution of Civil Engineers*, vol. 156, no. 3, pp. 137-41, Aug. 2003.

[15] M. Graff, "Learning from web-based instructional systems and cognitive style," *British J. of Educational Technology,* vol. 34, pp. 407-418, 2003.

[16] G. M. Johnson, "Learning style under two web-based study conditions," *Educational Psychology,* vol. 27, pp. 617-634, 2007.

[17] G. M. Johnson, "Student alienation, academic achievement, and WebCT use," *Educational Technology and Society,* vol. 8, pp. 179-189, 2005.

[18] C. P. Cadieux, "Variables associated with a sense of classroom community and academic persistence in an urban community college online setting," Ph.D. dissertation, Urban Services, Old Dominion Univ., Norfolk, VA, 2002, p. 1. (ERIC Document Reproduction Service No. Ed474545).

[19] D. McConnell, "Examining the dynamics of networked e-learning groups and communities," *Studies in Higher Education,* vol. 30, pp. 25-42, 2005.

[20] G. M. Johnson, "Perception of classroom climate, use of WebCT, and academic achievement," *J. of Computing in Higher Education,* vol. 17, pp. 25-46, 2006.

[21] K. L. Kelly and J. Schorger, *Online learning: Personalities, Preferences and perceptions*. U.S. Department of Education. 2002. (ERIC Document Reproduction Service No. ED470663).

[22] C. J. MacGregor, "Personality differences between online and face-to-face students," *J. of Continuing Higher Education,* vol. 50, pp. 14-23, 2002.

[23] I. B. Myers, *Introduction to Type*. Palo Alto, CA: Consulting Psychologists Press, 1993, p. 4.

[24] C. Soles and L. Moller. (2002). Myers Briggs type preferences in distance learning education. *International Journal of Educational Technology* [Online]. *2*. p. 42. Available: http://www.ascilite.org.au/ajet/ijet/v2n2/soles/index.html

[25] R. Daughenbaugh, D. Daughenbaugh, D. Surry, and M. Islam. (2002). Personality type and online versus in-class course satisfaction. *Educause Quarterly* [Online]. *3*. pp. 71-72. p. 72 Available: http://net.educause.edu/ir/library/pdf/EQM02312.pdf.

[26] J. W. Keefe, "Learning style: An overview," in Student Learning Styles: Diagnosing and Prescribing Programs, J. W. Keefe, Ed. Reston, VA: National Association of Secondary School Principals, 1979, pp. 1-17, p. 4.

[27] C. M. Karuppan, "Web-based teaching materials: A user's profile," *Internet Res.,* vol. 11, pp. 138–148, 2001.

[28] K. Sabry and L. Baldwin, "Web-based learning interaction and learning styles," *British J. of Educational Technology,* vol. 34, pp. 443–454, 2003.

[29] S. R. Terrell, "The effect of learning style on doctoral course completion in a Web-based learning environment," *Internet and Higher Education,* vol. 5, pp. 345–352, 2002.

[30] B. Bajraktarevic, W. Hall, and P. Fullick. (2003). Incorporating learning styles in hypermedia environment: Empirical evaluation [Online]. Available: http://wwwis.win.tue.nl/ah2003/proceedings/paper4.pdf.

[31] B. Mehlenbacher, C. R. Miller, D. Covington, and J. S. Larsen, "Active and interactive learning online: A comparison of Web-based and conventional writing classes," *IEEE Trans. on Professional Communication,* vol. 43, pp. 166-84, 2000.

[32] S. Y. Chen, G. D. Magoulas, and D. Dimakopoulos, "A flexible interface design for web directories to accommodate different cognitive styles," *J. of the American Society for Information Sci. and Technology,* vol. 56, pp. 70-83, p. 71, 2005.

[33] K. Loomis. (2000). Learning style and asynchronous learning: Comparing the LASSI Model to classroom performance. *Journal of Asynchronous Learning Networked* [Online]. *4(1)*. Available: http://www.aln.org/publications/jaln/v4n1/v4n1_loomis.asp.

[34] K. S. Kim and J. L. Moore. (2005, November). Web-based learning: Factors affecting students' satisfaction and learning experience. *First Monday* [Online]. *10(11)*. Available: http://www.firstmonday.org/issues/issue10_11/kim/index.html.

[35] G. M. Johnson, and A. Kupla, "Dimensions of online behavior: Toward a user typology," *CyberPsychology and Behavior,* vol. 10, pp. 773-780, 2007.

[36] A. N. Joinson, *Understanding the psychology of Internet behaviour: Virtual worlds, real lives*. New York: Palgrave MacMillan, 2003.

[37] S. C. Herring, "Computer-mediated discourse analysis: An approach to researching online communities," in *Designing for virtual communities in the service of learning*, S. A. Barab, R. Kling, and J. H. Gray, Eds. Cambridge, MA: Cambridge University Press, 2004, pp. 338-376.

[38] N. H. Nie, D. S. Hillygus, and E. Lutz, "Internet use, interpersonal relations and sociability: A time diary study," in *The Internet in Everyday Life*, B. Wellman and C. Haythornthwaite, Eds. Oxford, UK: Blackwell, 2002, pp. 215-243.

[39] S. Rodgers and K. M. Sheldon, "An improved way to characterize Internet users," *J. of Advertising Res.*, vol. 42, no. 5, pp. 85-94, Sep./Oct. 2002.

[40] R. Agarwal, A. Animesh, and K. Prasad. (2005, August). Social interactions and the digital divide: Explaining regional variations in Internet use. *Robert H. Smith School Res. Paper No. RHS 06-024* [Online]. Available: http://ssrn.com/abstract=796090

[41] I. M. Wasserman and M. Richmond-Abbott, "Gender and the Internet: Causes of variation in access, level, and scope of use," *Social Sci. Quarterly*, vol. 86, no. 1, pp. 252-270, Mar. 2005.

[42] J. Everett and R. E. Rice, *Social consequences of Internet use: Access, involvement, and interaction*. Cambridge, MA: MIT Press, 2002.

[43] S. Molnár, "Sociability and the Internet," *Review of Sociology of the Hungarian Sociological Association*, vol. 10, no. 2, pp. 67-84, Nov. 2004.

[44] G. Pronovost, "The Internet and time displacement: A Canadian perspective," *Instructional Technology and Society*, vol. 1, no. 2, pp. 44-53, 2002.

[45] J. Waldvogel. (2001, December). Email and workplace communication: A literature review. *Language in the Workplace Occasional Papers* [Online]. *3*. Available: http://www.vuw.ac.nz/lals/research/lwp/docs/ops/op3.htm.

[46] M. A. Fenichel, "Online behavior, communication, and experience," in *Online Counseling: A Handbook for Mental Health Professionals*, R. Kraus, J. Zack, and G. Stricker, Eds. Boston, MA: Academic, 2004, pp. 3-18.

[47] S. T. Kerr, "Why we all want it to work: Towards a culturally based model for technology and educational change," *British J. of Educational Technology*, vol. 36, no. 6, pp. 1005-1016, Nov. 2005.

[48] R. S. Nickerson, "Technology and cognitive amplification," in *Intelligence and Technology: The Impact of Tools on the Nature and Development of Human Abilities,* R. J. Sternberg and D. D. Preiss, Eds. Mahwah, NJ: Lawrence Erlbaum, 2005, pp. 3-27.

[49] Z. Papacharissi and A. M. Rubin, "Predictors of Internet use," *Journal of Broadcasting and Electrical Media*, vol. 44, no. 2, pp. 175-196, p. 178, Spr. 2000.

[50] M. Q. Huynh, J. N. Lee, and B. A. Schuldt, "The insiders' perspectives: A focus group study on gender issues in a computer-supported collaborative learning environment," *J. of Information Technology Education*, vol. 4, pp. 236-255, 2005.

[51] F. Volk and F. B. Kraft, "Consumers: The theoretical foundations of online behavior," in *Handbook of Human Factors in Web Design*, R. W. Proctor and K. L. Vu, Eds. Mahwah, NJ: Lawrence Erlbaum, 2005, pp. 595-612.

[52] G. M. Johnson, "College student Internet use: Convenience and amusement," *Canadian J. of Learning and Technology,* vol. 33, pp. 141-157, 2007.

[53] J. Kalbach. (2000, December). Designing for information foragers: A behavioral model for information seeking on the World Wide Web. *Internetworking* [Online]. *3(3)*. Available: http://www.internettg.org/newsletter/dec00/article_information_foragers.html.

[54] D. E. Mastro, M. S. Eastin, and R. Tamborini, "Internet search behaviors and mood alterations: A selective exposure approach," *Media Psychology*, vol. 4, no. 2, pp. 157-172, May 2002.

[55] A. Dix, "Human-computer interaction and web design," in *Handbook of Human Factors in Web Design*, R. W. Proctor and K. L. Vu, Eds. Mahwah, NJ: Lawrence Erlbaum, 2005, pp. 28-48.

[56] G. M. Johnson, "Learning style under two web-based study conditions," *Educational Psychology*, vol. 27, no. 5, pp. 617-634, Oct. 2007.

[57] M. Klusch, E. Andre, T. Rist, T. Ishida, and H. Nakanishi, "Interactive information agents and interfaces," in *Handbook of Human Factors in Web Design,* R. W. Proctor and K. L. Vu, Eds. Mahwah, NJ: Lawrence Erlbaum, 2005, pp. 221-236.

[58] K. Squire, "From content to context: Videogames as designed experience," *Educational Researcher*, vol. 35, no. 8, pp. 19-29, 2006.

[59] G. M. Johnson and J. A. Johnson, "Personality, Internet experience, and e-communication preference," in *Proc. of the IADIS International Conference WWW/Internet 2006, Vol. 1*, P. Isaias, M. B. Nunes, and I. J. Martinez, Eds. Murcia, Spain: IADIS, 2006, pp. 55-62.

[60] R. M. Felder and R. Brent, "Understanding student differences," *J. of Eng. Education*, vol. 94, no. 1, pp. 57-72, 2005.

[61] S. MacGowan, *The ultimate online dating guide*. Victoria, BC: Trafford, 2004.

[62] R. LaRose and M. S. Eastin, "A social cognitive theory of Internet uses and gratifications: Toward a new model of media attendance," *J. of Broadcasting and Electronic Media*, vol. 48, no. 3, pp. 358-377, p. 367, Sep. 2004.

[63] L. F. Capretz, "Clues on Software Engineers' learning styles," *International J. of Computing & Information Sci.*, vol. 4, no. 1, pp. 46-49, Apr. 2006.

Concept Classification Model for Digital Divide

J. Naarmala[1], O. Mäkinen[2]
[1]University of Vaasa, P.O. Box 700, FI-65101 Vaasa, Finland, email: jyri.naarmala@iki.fi
[2]Seinäjoki University of Applied Sciences, P.O. Box 412, FI-60101 Seinäjoki, Finland, email: olli.makinen@seamk.fi

Abstract – **Digital divide has been used as a central concept in number of studies, yet this concept has many different definitions. This paper discusses the meaning and definition of digital divide and provides a dichotomy to be used for analyzing different definitions of this concept. Main contribution of this paper is on providing a theoretical dichotomy for understanding the meaning and underlying assumptions related to concept of digital divide.**

Keywords – **Digital divide, Knowledge Society, Conceptual analytics.**

I. INTRODUCTION

Contemporary western society has undergone significant changes, while giant leaps in technological development have made it possible for the majority of western world countries to enjoy the benefits of knowledge society. While evolving from post industrial society (a.k.a information society) towards knowledge society, the way how computer networks are utilized by society and individuals for providing and/or accessing digital services has changed. While public digital services were not that common fifteen years ago, those are an inseparable part of contemporary society. Those who can not access, nor use these services are slowly being isolated from majority of services, while those very same services are moving increasingly online.

These outcasts of information society are often referred as being victims of digital divide, while the definition what digital divide really implies appears to vary depending on the context where it is being used. Usually it is referring to restricted or total lack of accessibility to online services. Reasons for this might vary from computer literacy to totally missing infrastructure.

Norris [1], for instance, raises into focus the fact that the global digital divide is substantial and has been growing during the first decade of the internet age. Payton, on the other hand, states that [2], while the efforts to eliminate the digital divide are valuable, public policymakers must also seek a way to improve the social network that will help the digital divide to resolve itself.

According to Norris [3] digital divide has many faces. Norris asks if the fast and revolutionary development of information technology will reinforce or erode the gap between information-rich and poor nations, if it will exacerbate or reduce social divisions within different countries or strengthen representative democracy or, on the contrary, buttress the power of established interests, both the political and economical structures. So, the question is whether information technology is a restructuring or the maintaining force.

This theoretical paper focuses on developing a theoretical tool to be used, for analyzing different interpretations of digital divide on a conceptual level. The meaning of this concept is both vague and obscure. Could be said that the significance of this concept has not vanished, but its inner meaning is starting to be outdated and is in a need for thorough revision. The main goal of this paper is to present existing definitions of digital divide and to show how the concept should be updated. As such, this paper exercises approach commonly used in conceptual analytical research [4]. The path of the conceptual analysis is as follows: first the basic assumptions behind constructs are analyzed. Then theories, models and frameworks used in empirical studies are identified. Finally logical reasoning is applied. [5]

Following the footsteps of Burrell and Morgan [6] and Deetz [7], the goal is to develop concepts suitable for making classifications, while focusing on large scale phenomenon, digital divide.

Many definitions concerning digital divide are focusing either on technological or social issues, while some are focusing on both. Because the concept appears to be evolving a classification to be later refined into taxonomy is necessary.

II. METHOD

This study is theory creating study [8], presenting a conceptual-analytical study. In this study concepts derived from IT impact literature are further developed and refined to be used in analyzing digital divide phenomenon. Concepts are used for defining polarities, which are used for constructing a theoretical framework to be used for concept classification. In this paper authors are entwining together concepts from social sciences and information systems (IS) literature, where the goal is to be able to combine technical and social viewpoint on endeavour for better understanding global phenomenon. Used approach in this conceptual-analytical study is induction, where through generalizing contributions from previous studies and enhancing the concepts, a new approach can be presented.

Lee & Baskerville [9] suggest, that when generalizing concepts from theory, a researcher generalizes from theoretical propositions in the form of concepts to the theoretical propositions that make up a theory (specifically, a set of logically consistent propositions that, pending the

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_14,

results of empirical testing, could qualify as a theory). In addition to this, another form of generalizing from concepts to theory is the formulation of a theory based on the synthesis of ideas from a literature review [10]. In this paper, concepts from IT impact literature are adopted and enhanced to suit the use in different context.

III. EVOLVING PARADIGMS

In 1979 Burrell and Morgan introduced four paradigms for social science research. Presented theoretical construction was based on two dimensions, which were subjective — objective and regulation — radical change. These paradigms should be seen as contiguous, but separate; contiguous because of the shared characteristics, and separate because of the differentiation. They state that these four paradigms define perspectives for the analysis of social phenomena, which are fundamentally different. [11]

Although widely used, Burrell and Morgan's model is in some ways insufficient. In 1996 Deetz introduced a model that both criticizes and extends legacy of Burrell and Morgan. Deetz presented critique on subjective-objective problem in Burrell and Morgan's model. Deetz presented three arguments: 1) the meaning of the objective-subjective labels is already socially contrived and Deetz treats the claim of objectivity or subjectivity as a rhetorical move in research program's system of justification rather than as a useful descriptive label. According to him neither all research is both subjective and objective. He also suggests that subjectivity and objectivity just are not very interesting ways of thinking about research program differences; 2) The subjective-objective conception reproduces a neo-positivist philosophy of science and obscures the nature of other research programs instead of describing a meaningful difference; 3) Continuation of conflicts and misleading presumed relations between so-called qualitative and quantitative research, due to the conception of subject-object separation. [12]

Deetz presents two new dimensions of contrast. First dimension focuses on the origin of concepts and problem statements as part of the constructive process in research, and the second dimension focuses on the relation of research practices to the dominant social discourses within the studied organization, the research community, and/or wider community. Differences among research orientations can be shown by contrasting "local/emergent" research conceptions with "elite/a priori" ones, while differences among the relation of research practices to the dominant social discourses can be shown by contrasting "dissensus" with "consensus" as illustrated in Fig. 1. [13]

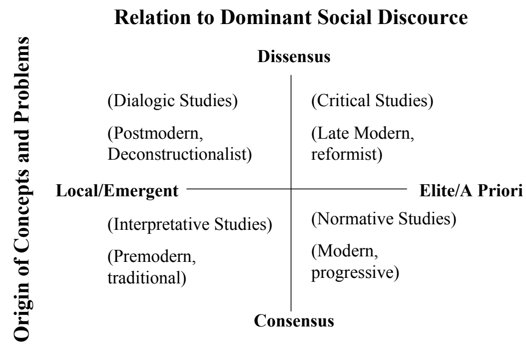


Fig. 1. Contrasting dimensions from the metatheory of representational practices [14].

Orlikowski [15] introduced interpretative flexibility of technology in her study about different perceptions of technology in organizations. Interpretative flexibility can be defined as "*the capacity of a specific technology to sustain divergent opinions*" [16]. Interpretive flexibility has long been recognized as playing an important role in explaining how technical artefacts are socially constructed [17], thus being a suitable concept to be used in the context of the framework of this study. In this paper this concept is partially refined, and according to this new definition this concept has two polarities: challenging technology and immersive technology. These concepts imply the subjective perception of adjustability, adoptability and usefulness of chosen technology, varying from challenging to immersive.

Immersive virtual reality is a hypothetical future technology that exists today as virtual reality art projects, for the most part. It consists of immersion (virtual reality) in an artificial environment where the user feels just as immersed as they usually feel in consensus reality. The technology is immersive when it wraps its users in its safe embrace.

In many digital divide definitions, the concept has also social dimension (while using demographic variables as explaining factors, such as age, cost, family structure, religion, etc.) in addition to technological one. Therefore in this paper another concept to be used is social dimension, and like interpretative flexibility of technology, also this has two polarities. These are defined in this paper as social inclusion and social exclusion, where first implies that demographic and social issues create a "pull effect", which in turn keeps people within the domain of knowledge society, while the latter implies the opposite. Fig. 2 illustrates these dimensions, where y-axis illustrates the interpretative flexibility of technology and x-axis the social dimension.

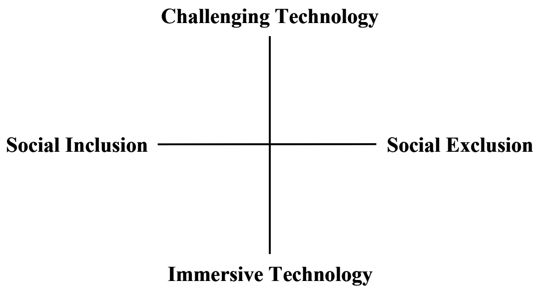


Fig. 2. Dichotomy of social-technical dimensions.

When utilizing model presented in Fig. 2, we end up with four different categories, or definitions. In general, it appears that this dichotomy suggests how concepts should be further developed by suggesting what is lacking in definition. While combination Social Exclusion - Challenging Technology appears to be the worst case scenario and Social Inclusion - Immersive Technology is the best, there are definitions that are not clearly just one or another, but something in between. This model provides simple guidelines and target, where to set goals while further developing concept definitions. Ideal models presented in Fig. 2 are discussed in detail below.

A. Challenging Technology – Social Inclusion

Here technology is seen as a challenge, and people are not sure about their computer skills or the available online services in general. The technology is perceived as immature or it is not available for everyone. Social inclusion implies that these people are economically secured, active members of society and active parts of social networks.

In this situation the IT development is too fast. The citizens are not able to adopt all the new innovations. But there is another interesting group of the non-users of the Internet. These people, reluctant to use the Internet, maybe can afford to use the Internet and to acquire the necessary equipment, but they appreciate the traditional ways of communication and handling of their affairs because in that way they are able to maintain the social contacts with the suppliers of the services. In this respect they belong to the group of the "voluntarily nonusers".

That is to say, some people have voluntarily decided to stay outside the information networks. Therefore we can't speak of digital divide, because it is always possible to presuppose that the people are able to make a connection to the Internet whenever they want. This is, however, a false conclusion or at least a generalization. The real reason may be any of the above-mentioned. Maybe people are not able to use the Internet because of the lack of technical skills. They could have possibilities to acquire education or guidance. Why don't they do it? It is possible that they are ignorant – they are not interested in the contents of the Internet. Or they estimate that the efforts to catch up the skills are too laborious.

B. Challenging Technology – Social Exclusion:

Here technology is seen as challenge. In addition to this social exclusion refers to marginal groups (which are not necessary always that marginal) which are either totally outside from the safe domain of knowledge society, whether based on the socio-economical status, or lacking computer literacy, or broadband access. This is the worst type of digital divide, because these people are separated from the knowledge society in two ways.

The traditional victims of digital divide belong to this category. The usual definitions of digital divide point out for instance the social factors and stress the equality: "[digital divide is] *a division between those in favor of the extensive use of digital technology (esp. computers) and those against it; (b) (now the usual sense) the gulf between those who have ready access to current digital technology (esp. computers and the Internet) and those who do not; (also) the perceived social or educational inequality resulting from this* [---]" So, the concept is used here to refer in the first place to unequal access to computers and computer networks. [18]

C. Immersive Technology – Social Inclusion

Here technology is seen as a enabling and supporting factor that is useful, easily adopted and adjustable according to individual preferences. Social dimension is seen as supporting the integration of these people into knowledge society by guiding them to information resources and services available in knowledge society.

This kind of society is Habermasian, that is to say both utopian and communicative. According to Jürgen Habermas, the function of philosophy is to promote understanding and democracy. He says in his *Theory of Communicative Action* [19] that there is a communicative void in the society. The more people and institutions communicate in a society, the more efficiently democracy works [20]. According to this reading, Habermas and the Critical Theory are not neutral and objective, but speak out. This is certainly true, as the purpose of the Critical Theory and Theory of Communicative Action is to improve positive values including freedom, consensus, communicative skills, and democracy. Is such a theory, then, objective? Could it not be called a theory of normative ethics? Here digital divide does not exist.

D. Immersive Technology – Social Exclusion

Technology is seen as a enabling and supporting factor, while social exclusion is seen as isolating people from available access to information resources and services. Social dimension is seen as hindering factor that translates into lacking social networks, poor economical status, illiteracy, or lacking computer skills. Generally speaking in these cases the technology is present, but social barriers hinder access to these information resources.

In this situation the society would invest in information literacy education to all age groups, as well to young people as senior citizens. According to the information literacy standards launched Association of College and Research Libraries the education of information literacy makes it possible to make good use of the huge amount information and the highly developed information channels we have today [21]. The free accessibility to all these information sources is also a democratic question. [22] Edification and information literacy education is also important for the citizens from a legal protection standpoint; we see more and more public services online as time advances and especially the elderly and the handicapped no doubt are in danger of ending up in the digital gap. Guaranteeing them service (knowledge, instruments) is an ethical question. [23]

IV. USING THE MODEL FOR ANALYSIS - THE CUSTOMARY MEANING OF THE CONCEPT OF THE DIGITAL DIVIDE

Now it is possible to test our model and arrange different kind of definitions to it. Many scholars have been worried about the inadequate telecommunication skills of aging adults. The social services i.e. the service culture is becoming more and more digitalized. Vesa Mäensivu [24] points out that the reasons for low communication in digital environments are manyfold. Partially the aged are afraid that they might not have the necessary learning ability to use the new equipment.

Mäensivu names a long list of factors or obstacles that prevent the elder from using the Internet. Among those are the obstacles for accessibility i.e. the traditional reasons of the digital divide. They can be divided into technical and human factors [25]. The Web pages are often considered complicated and difficult to use. The elder often have impaired vision and they feel the text size on Web pages is quite too small. Some background colours, red and green, were mixed together. Logging on to Internet services was too difficult to some of the group Mäensivu had interviewed. They were embarrassed when they did not know what to write in the login box. The use of a computer mouse – for instance the difference between the click and the double click – was one of the basic obstacles [26]. It's quite clear that the use of the Internet becomes almost impossible if you can't control the mouse. Some members of Mäensivu's group had it difficult to distinguish the difference between the click of the mouse and the press of enter. These kind of obstacles obviously belong to Group A: *Challenging Technology – Social Inclusion.* The citizens can afford to the new technology but they feel it too complicated.

The human factors, as the above mentioned fear and anxiety about one's own IT skills, offer one interesting point of view. One of the main obstacles to the use of the Internet is traditionalism. Senior citizens think they don't need the Internet services if they are able to look after things and manage the finances in traditional ways. Two thirds of the aged Mäensivu interviewed agreed completely with the statement that insisted on personal, face to face contacts when handling their affairs. [27]

It is apparent that the digital divide has many faces. Traditionalism in the attitudes of the senior citizens may have nothing to do with digital divide as seen before. These people, reluctant to use the Internet, maybe can afford to use the Internet and to acquire the necessary equipment, but they appreciate the traditional ways of communication and handling of their affairs because in that way they are able to maintain the social contacts with the suppliers of the services. In this respect they belong to the group of the "voluntarily nonusers". People in this category may also belong to the victims defined in the group B *Challenging Technology – Social Exclusion.*

According to lexical definition the digital divide refers to: "*[Digital divide is] a division between those in favor of the extensive use of digital technology (esp. computers) and those against it; (b) (now the usual sense) the gulf between those who have ready access to current digital technology (esp. computers and the Internet) and those who do not; (also) the perceived social or educational inequality resulting from this [---]*". The concept is used here to refer in the first place to unequal access to computers and computer networks. It is argued here that the original definition of digital divide does need a revision, because the nature of division is no longer simply based on the division between those who favor technology and those who do not, nor on the access to digital technology. Here the society is often unequal and the people in the gap also belong to the category B: Challenging Technology – Social Exclusion.

According to Economist Intelligence Unit e-readiness ranking [28] a critical shift is occurring in the social fabric of countries: as governments and enterprises increasingly migrate services online, those without reliable Internet access are in danger of being disenfranchised, e.g. the concrete risk for digital divide is present. Due to this both public- and private-sector organizations have spent considerable time and money on trying to get economically disadvantaged communities online. Increasingly, these attempts are spreading to the socially disadvantaged as well. According to report [29] senior-citizen access has been a particular focus for developed countries, where some 20,000 senior citizens are estimated to have been trained through various initiatives. EIU report (2007) seems to combine concepts of economical and social disadvantages as digital divide, whereas technical ability is not seen as a key factor, like in OED definition. People suffering from digital divide as defined above belong to the group D: *Immersive Technology – Social Exclusion*

Norris [30] raised into focus a fact that the global digital divide is substantial and has been growing during the first decade of the internet age, and Payton stated that [31] while the efforts to eliminate the digital divide are valuable, public policymakers must also seek a way to improve the social network that will help the digital divide to resolve itself.

According to Norris [32] digital divide has many faces. Norris asks if the fast and revolutionary development of information technology will reinforce or erode the gap between information-rich and poor nations, if it will exacerbate or reduce social divisions within different countries or strengthen representative democracy or, on the contrary, buttress the power of established interests, both the political and economical structures. So, the question is whether information technology is a restructuring or the maintaining force. The definitions of Pippa Norris describe well the digital gap as it is defined in group B: *Challenging Technology – Social Exclusion.*

It is said that the Internet places different demands on the user than the medium of television-access and its use requires both cognitive ability and technical skills [33] [34] Most Finnish and Swedish people have "reasonable" access to the Internet. There is, however, a constant amount of people who voluntarily or involuntarily have chosen to stay outside the information networks. What is the reason for that? It can't be said that in the rich Nordic countries the cause of digital divide could be only ethnical or depend on factors like age, costs, family structure, income, geography, rural and regional factors, religion, education and so on. So this kind of digital divide is a good example of the group D: *Immersive Technology – Social Exclusion*

Some people have voluntarily decided to stay outside the information networks; they also belong to the group D. Actually we can't speak of digital divide, because it is always possible to presuppose that the people are able to make a connection to the Internet whenever they want. This is, however, a false conclusion or at least a generalization. The real reason may be any of the above-mentioned. Maybe people are not able to use the Internet because of the lack of technical skills. They could have possibilities to acquire education or guidance. Why don't they do it? It is possible that they are ignorant – they are not interested in the contents of the Internet. Or they estimate that the efforts to catch up the skills are too laborious. In our model these non-users are still "citizens" in knowledge society, although they are staying aside from the main stream.

V. CONCLUSIONS AND FURTHER RESEARCH

Presented model works as a starting point for developing taxonomy for digital divide concept classifications. While being an issue of great importance for evolving western societies, these very basic questions have not been thoroughly defined. Digital divide is not just about technology, nor social issues. It is a combination of these both, a mixture that is extremely challenging. As always, to be able to resolve problems, it is necessary to ask right questions. But how can these right questions be asked, if the very basic foundations have not been properly defined?

It is clear that digital divide is, or defining the meaning of (the concept) digital divide is a cross-disciplinary project. The Dichotomy of social-technical dimensions that the authors have developed helps to understand the phenomenon of the digital divide that can be ethical and democratic question.

Equality is also an important regulative principle or an aiming point in this research. It is not "unscientific" to say, that digital divide should be avoided. Digital divide refers to a society that is not working effectively, as we will see later on in this study. This starting point this study isn't however to support the normative ethics – that standpoint will we pass over this time [35]

The society is becoming all the time more complicated. The technical, economical and "scientific" crème de la crème have the power. Democracy is only apparent. According to the communication theory a model of democratization based on the domination-free communication that should first take place among scholars could be the last possibility, the salvation [36]. Could the Internet be the basis for a fuller and richer forum for political debate? Could it develop into the kind of public space which Jürgen Habermas, for instance, advocated for a genuine democracy? [37] If this should come true, the democracy within the Internet should be first realized. The digital divide should not exist. Habermas speaks about mutual understanding or consensus [38]. According to Kotkavirta [39] intersubjectivity is one of the main themes in the communication theory of Jürgen Habermas. In spite of the breech between Habermas and the traditional standpoints of the Critical Theory, Habermas was still stressing on the criticism on the society or the culture as the emancipatory task of philosophy and Social Science [40].

REFERENCES

[1] Norris, P. (2001). *Digital Divide – Civic Engagement, Information Poverty, and the Internet Worldwide.* Cambridge University press: USA, p.233.
[2] Payton, F. C. (2003). Rethinking the digital divide. *Communications of the ACM*, Vol. 46 Iss. 6, pp. 89-91.
[3] Norris, P. (2001), pp 3, 95-112.
[4] Järvinen, P. (2004) On Research methods. Opinpajan kirja: Tampere, p. 17.
[5] Ibid, p. 19.
[6] Burrell, G. & G. Morgan (1988). Sociological Paradigms and Organizational Analysis. England: Gower publishing Company Limited. Reprint.
[7] Deetz S. (1996), Describing differences in approaches to organization science: Rethinking Burrell and Morgan and their legacy, Organization Science 7, No 2, 191-207.
[8] Järvinen, P. (2004) On Research methods. Opinpajan kirja: Tampere.
[9] Lee, A.S. & Baskerville, R.L., 2003. Generalizing Generalizability in Information Systems Research. Information Systems Research, Vol. 14, No. 3, p. 238.
[10] ibid.
[11] Burrell, G. & G. Morgan (1988). Sociological Paradigms and Organizational Analysis. England: Gower publishing Company Limited. Reprint. p. 21-23.

[12] Deetz S. (1996), Describing differences in approaches to organization science: Rethinking Burrell and Morgan and their legacy, Organization Science 7, No 2, p. 193-194.
[13] Ibid., p. 195.
[14] Ibid., p. 198.
[15] Orlikowski, W.J. (1992). The duality of technology: Rethinking the concept of technology in organizations. Organization Science, 3:3, 398-427.
[16] Sahay S. & D. Robey (1996). Organizational context, social interpretation, and the implementation and consequences of geographic information systems. Accounting, Management and Information Technology, 6:4, p. 260.
[17] Doherty, N.F., Coombs, C.R. & J. Loan-Clarke (2006). A re-conceptualization of the interpretive flexibility of information technologies: Redressing the balance between the social and the technical. European Journal of Information Systems, 15:6, 569-582.
[18] OED Online, cited October 13th, 2008.
[19] Habermas, J. (1984). *The Theory of Communicative Action.* (T. McCarthy, Ed.). (Vol. 1). Boston: Beacon Press.
[20] Salter, L. (2003). Democracy, New Social Movements, and the Internet: A Habermasian Analysis. In L. L. Brennan & V. E. Johnsson (Eds.), *Cyberactivism: Online Activism in Theory and Practice.* (pp. 117-144). New York: Routledge.
[21] Association of College and Research Libraries (2001). *Information Literacy Competency Standards for Higher Education.* Retrieved October 13th, 2008, from http://www.ala.org/ala/acrl/acrlstandards/informationliteracycompetency.cfm
[22] Holmlund, S., Mikola, M., Mäkinen, O. (2008). Good and Bad in Information Retrieval – Ethics in Information Literacy Education: Case *The Tritonia Academic Library*. Vaasa: Tritonia (IN PRINT).
[23] Ibid.
[24] Mäensivu, V. (2003). *Ikääntyvien viestintävalmiudet ja digitaalinen epätasa-arvo* (*Telecommunication skills of aging adults and the risk of a digital divide*). Helsinki: KELA, pp. 158-161.
[25] Ibid, 159.
[26] Ibid, p. 160.
[27] ibid.
[28] Economist Intelligence Unit. (2007). *The 2007 e-readiness rankings, Raising the bar*. A white paper from the Economist Intelligence Unit. Cited October 5th, 2008. Retrieved from: http://graphics.eiu.com/files/ad_pdfs/2007Ereadiness_Ranking_WP.pdf
[29] Ibid.
[30] Norris, P. (2001). *Digital Divide – Civic Engagement, Information Poverty, and the Internet Worldwide.* Cambridge University press: USA, p.233.
[31] Payton, F. C. (2003). Rethinking the digital divide. *Communications of the ACM*, Vol. 46 Iss. 6, pp. 89-91.
[32] Norris, P. (2001), pp. 95-112.
[33] Curtin, J. (2001). *A Digital Divide in Rural and Regional Australia?* Cited October 5th, 2008. Retrieved from: http://www.aph.gov.au/LIBRARY/pubs/cib/2001-02/02cib01.htm
[34] Bucy, E.P. (2000). Social Access to the Internet. *Press/Politics* 5 (1), pp 50-61.
[35] von Wright, G. H. (1968). *The Varieties of Goodness.* London: Routledge & Keagan Paul, pp. 3-5.
[36] Wiggershaus, R. (1994). *The Frankfurt School. Its History, Theories and Political Significance.* Cambridge: Polity Press, p. 652.
[37] Ferdinand, P. (2005). Cyberpower: Only the power to Disturb? *Concilium* 2005/1, pp. 27-35.
[38] Habermas, J. (1994). Kommunikatiivisen toiminnan käsitteen tarkastelu. In Jürgen Habermas *Järki ja kommuniikatio* (ed. and transl. by Jussi Kotkavirta). Helsinki: Gaudeamus, pp. 68-97.
[39] Kotkavirta, J. (1994). Jälkisanat. In In Jürgen Habermas *Järki ja kommuniikatio* (Ed. and transl. by Jussi Kotkavirta). Helsinki: Gaudeamus, pp. 240-260.
[40] Ibid, 242.

A Virtual Explosion or SNAFU is *Always* Better Than a Real One: Exploring the Use of Virtual Worlds for Simulation and Training…and Developing the Leaders of Tomorrow

David C. Wyld
Southeastern Louisiana University
Department of Management – Box 10350
Hammond, LA 70402-0350

Abstract-**This paper explores the growing use of virtual world simulations for training and education, both in the private and public sectors. We examine the rise of the importance of the rise of the "digital native" generation and the context of their growing comfort operating in online gaming and increasingly, in virtual worlds. We look at the growth of the video game and virtual world marketplaces and the demographics of players in these environments, which are largely populated by members of the digital native generation. We then examine how leading-edge organizations, from the U.S. Army to IBM to public agencies and universities training emergency responders are making use of virtual worlds for not just conducting simulations and training in cost and time-effective – not to mention safer – manners. We also look at emerging research that demonstrates how virtual worlds can be used to not develop field leadership and teamwork skills key to success in today's organizations. We conclude with a discussion on the importance to fit the needs for today's digital natives and the desires for improved training and education.**

I. INTRODUCTION

A. *Video Game Nation*

Video games have become a huge part of not just the entertainment marketplace, but of life today. Anyone who has kids can testify to the power of video games to entrance their child's attention for hours (and hours) upon end. And the data support our casual observances of the crowds around Wii's and playing Guitar Hero, RockBand, and thousands of other games. And it is not just the pre-teen set. On college campuses, in coffee houses, in airports, and yes, in office cubicles, people of all ages are increasingly playing video games. All this activity adds up to a huge market. According to recent estimates from market analyst DFC Intelligence, the video game industry exceeded a $33 billion marketplace in 2006 – and it is expected to reach $47 billion by 2009 [1].

The latest survey of the U.S. video game marketplace shows almost three-quarters of all Americans play video games of one type or another, with well over half of these not just playing games on their personal computers, game consoles, or increasingly, their mobile devices, but online as well [2]. Online gaming is exploding in popularity, and the largest growth of online gaming is amongst the young. And the younger the kid, the more likely it is that they will be online gamers. According to the 2007 American Kids Survey, nearly 4 in 5 American kids between the ages of 6 and 11 play online games [3]. This has amounted to rapid growth of the Massively Multiplayer Online Games (MMOG) sector, with conservative estimates putting the total number adult subscribers at over 17 million today [4].

As can be seen in Figure 1, a great deal of this activity is not just with online video games, but in online virtual worlds. In fact, online gaming has become the most popular of all online activities for today's youth – far exceeding the percentage of pre-teens who do any other activity, including, of course, doing "stuff for school/homework" [5]. Anne Marie Kelly, Vice President of Marketing & Strategic Planning for the market research firm Mediamark Research and Intelligence, observed that the study's findings demonstrate that: "online gaming is clearly firmly entrenched as a pastime in the lives of most American kids" [3]. Indeed, while the percentage of the American population playing online video games decreases in older age groups, still, over half of all Baby Boomers and almost 4 in 10 members of the "mature generation" (those presently over the age of 61) play games online [2].

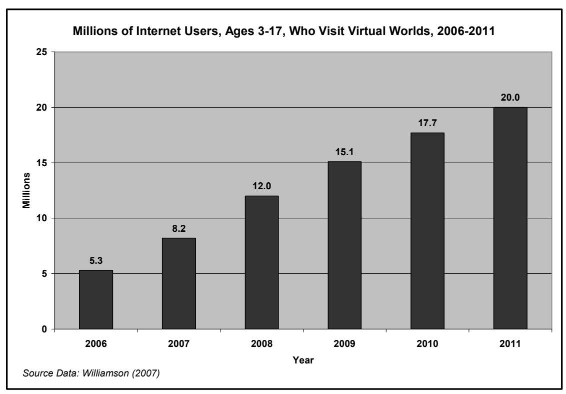


Fig. 1. The Growing Popularity of Virtual Worlds

Analysts have predicted that by 2020, "virtual worlds will be as widespread as the World Wide Web is now" [6], and there is a growing belief that virtual worlds may well "replace

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_15, © Springer Science+Business Media B.V. 2010

the web browser as the way we interface with the Internet" [7]. Indeed, some have predicted that virtual worlds will be as significant a technological disruptor as the invention of the personal computer or the advent of the Internet [8]. In late 2007, Gartner forecast that by the end of 2011, fully 80% of all active Internet users "will have a 'second life'" in the developing sphere of virtual worlds, "but not necessarily in Second Life" [9]. While some have criticized Gartner's 80% projection for being overly optimistic [10], there can be no doubt that the way we interact via the Internet is undergoing a profound change with the advent of virtual worlds.

Kid-oriented virtual worlds actually dominate the American marketplace. Big media and entertainment companies, like Walt Disney and Viacom's Nickelodeon division, are noting this trend, acquiring virtual worlds' sites aimed at kids (like Webkinz and Club Penguin) and creating virtual worlds for their theme parks, movies, and characters [11]. Kids – or their parents – buy a physical Webkinz stuffed animal, and then, they can go online and – using a code on the nametag of the animal – adopt and care for their pet in the Webkinz virtual world or buy and furnish an igloo in Club Penguin [12]. Likewise, toy companies, such as Mattel and Lego, are creating virtual worlds for children that correspond to their popular products. This burgeoning kids market has led analysts to predict: "Forget Second Life. The real virtual world gold rush centers on the grammar-school set" [13].

Observers have pronounced that "For young people, it will feel like second nature to live in a virtual world like Second Life" [14]. Studies have shown that on average, players log approximately twenty hours per week in virtual worlds [15]. Why do kids flock to virtual worlds? The answer may lie in that: "It seems that kids unlike adults are very good partners to the idea of alternative identities and cultures. I think kids perceive second worlds as games platforms which offer some more reality features. For adults second world's concept takes an opposite direction. It provides a visualized experience with almost no resemblance to your reality" [16] Weiss (2007).

Today's youth are thus coming to see virtual worlds as one of their primary ways to interact over the Internet. As they interact with virtual worlds, this changes their perspective on what their Internet experience can and should be. Virtual worlds, such as Second Life, are fast transforming the very definition of "being together." In these immersive environments of the "3D Internet," people from all over the country – indeed the world – can interact with one another in real-time. As such, virtual worlds have been labeled as a transformative development that will perhaps change the way we use the Internet over the next decade. Today's youth "are growing up not only with social networking but also with the ability to interact with people, shop, learn and play in a graphic environment" [5]. In comparison, "flat" websites may seem boring to the rising generation, paling in comparison to what they see as truly a Web 2.0 - a "new and improved" version of the Internet.

II. TRAINING AND LEARNING FOR "DIGITAL NATIVES"

Why are online games and virtual worlds such appealing concepts to both today's organizations and learners? The answer can be found in a generational quake that is occurring in real-time and being hotly discussed in the education literature today. Today's younger generation can be said to be "digital natives," a term coined by Marc Prensky to describe the rising generation as being "native speakers of technology, fluent in the digital language of computers, video games, and the Internet" [17]. All of us who are old enough not to be native to the digital culture are thus said to be digital immigrants. As such, while we may adopt many aspects of the technology, akin to those who learn a second language, "we retain an 'accent' because we still have one foot in the past" [17]. Today's parents, educators, political and corporate leaders are all largely digital immigrants, and as such, they can be said to "speak DSL, digital as a second language" [18].

Prensky conlcudes that as such, members of the younger generation "generally have a much better idea of what the future is bringing than we do" [19]. This is evidenced by the fact that digital natives are readily "busy adopting new systems for communicating (instant messaging), sharing (blogs), buying and selling (eBay), exchanging (peer-to-peer technology), creating (Flash), meeting (3D worlds), collecting (downloads), coordinating (wikis), evaluating (reputation systems), searching (Google), analyzing (SETI), reporting (camera phones), programming (modding), socializing (chat rooms), and even learning (Web surfing)" [19].

James Gee, who is a Professor of Reading at the University of Wisconsin-Madison and the author of *What Video Games Have to Teach Us about Learning and Literacy*, is adamant in calling video games "a much more powerful form of learning" than traditional classroom instruction. This is because "learning isn't about memorizing isolated facts. It's about connecting and manipulating them" [20]. He believes that today's new generation of learners are truly "learning' through playing casual video games – even violent and often-derided games such as *Grand Theft Auto: Vice City* and *Metal Gear Solid 2*. In game play, gamers are challenged to adapt and master new skills. Yet, too often, the same individuals are failing in the traditional educational realm, where too often, they are bored by the routine of the school environment [20].

Indeed, the very future of American training and education is at stake in the challenge to adapt the educational setting to the needs of digital natives. Prensky observed: "Pragmatically, our 21st century kids' education is quickly bifurcating. The formal half, 'school,' is becoming an increasingly moribund and irrelevant institution. Its only function for many students is to provide them with a credential that their parents say they need. The informal, exciting half of kids' education occurs 'after school.' This is the place where 21st century students learn about their world and prepare themselves for their 21st century lives" [19].

III. Serious Games

The use of simulations – both physical and computerized – has a long history, originating in military training. In fact, the use of simulations in warfighting can be traced back as far as the Roman Empire, with Roman commanders' use of "sand tables." These tables, which were a small copy of the physical battlefield, allowed for commanders to test their battle strategies using icons representing soldiers and units [21].

Still, Dr. Roger Smith (2008), the Chief Technology Officer for the U.S. Army Simulation, Training and Instrumentation Program Executive Office, advised policymakers to steer clear of any mention of the word "game" when speaking of simulations, as: "The word game is not appropriate in many communities" [22]. In fact, the U.S. military originally developed the term "serious game" as "a more acceptable way to talk about 'war games' with Congress and the public" [23].

Serious games are especially suited for not just everyday corporate and organizational settings, but for learning exercises for military personnel and first responders to practice responses to potentially catastrophic events in a safe environment. Such virtual reality simulations will be especially important as an alternative to live action training in life or death training environments and for conducting such exercises amongst remote participants – without travel or staging costs and the ability to literally "rerun" simulations and "tweak" exercises unlimited ways. Also, training in a virtual environment is inherently safer than engaging in a "real" exercise, as put it bluntly: "Every explosion that happens in a virtual world is a good thing. Every time a soldier is 'injured' in a simulation, they learn more about how to stay safe in reality" [24].

The virtual environment also allows for better in-game and post-hoc analysis of training exercises through the use of machinma video captures of the simulated experience. Machinma, which is a form of filmmaking that uses the capabilities of the virtual world technology to shoot films within the simulation, is exploding in use [25]. Robert Bowen, who is the civilian chief of the U.S. Army's simulated warfare games efforts, stated that in his realm, the goal is to able to "immerse that soldier into a virtual or synthetic environment, then have them conduct a training task, using their SOP [standard operating procedures], and then AAR [after-action review] that capability" [26]. Machinma make such AARs easy to make, record and use in the virtual environment for any form of training exercise or simulation – both military and civilian. Even perhaps more importantly, machinma can be used for analyzing "what if" situations for "virtual" terrorist actions, explosions, toxic waste spills, or industrial accidents. In fact, they may become an integral part of management and leadership-focused simulations held in virtual environments as well, making it possible to review social interactions and leadership exercises in new and exciting ways with the same after-exercise capabilities.

IV. Training in the Virtual World

Derryberry declared that: "21st century learning experiences need to reflect the lives of 21st century workers. While there is still skepticism that something called a game could be anything more than a leisure activity, serious organizations are getting serious results with serious games" [23]. Virtual worlds such as Second Life and World of Warcraft, as well as private virtual environments, are proving to be effective training and learning environments. This is because they provide what Lucia Graves of *U.S. News & World Report* recently described as a space that "provides a social laboratory where role-playing, simulations, exploration, and experimentation can be tried out in a relatively risk-free environment" [27]. In a game, an individual will try and fail — often many, many times — to learn the moves necessary to "slay the dragon" to either earn points and/or advance levels. Unlike in traditional training and learning environments, there is "permission to fail" in the gaming, as small failures are built into the process to allow for overall learning and advancement to take place [28]. Thus, urging organizations to make use of virtual games for training is no longer "techno-utopian silliness." As one observer remarked: "Making work more fun doesn't necessarily mean doing it in a multiplayer online game like World of Warcraft. But it does mean that businesses need to take some of the techniques games like Warcraft use to make repetitive work fun" [28].

We have seen pioneering examples today that will translate into models for wider use of serious games and virtual worlds-based training and simulations in the coming years. In fact, the U.S. Army has been a leader in using computer "games" as an active and reserve force-training tool since the 1980s – from tank simulations to combat mission simulators [29]. Today, the Army has an entire office dedicated to exploring the use of video games. At Fort Leavenworth, Kansas, as part of the Training and Doctrine Command's (TRADOC's) National Simulation Center, there exists the Project Office for Gaming, or TPO Gaming for short. The office was created in part out of the frustration that soldiers and units were often buying first-person shooter and military strategy games from Best Buy and other outlets – out of their own pockets – as a way of preparing for real-world situations. TPO Gaming is looking to create a simulation toolkit for creating realistic scenario training in virtual world, MMOG environments, with the target date for release between 2010 and 2015 [30]. The biggest challenge with military simulations, according to Colonel Jack Millar, who is the Director of TPO Gaming, is to create geospecific terrain that is accurate enough to present mission-specific rehearsal and training capabilities [31]. Brigadier General Thomas Maffey, who is the Army's Director of Training, based at the Pentagon, foresees a range of simulations for a variety of military training situations, stating: "While one game might provide excellent battlefield visualization, another might support training bilateral negotiation techniques" [31].

In the virtual world of Second Life, leading edge companies are actively experimenting with using the virtual world for real work. IBM is one of the leading companies seeking new and innovative ways to use virtual worlds in practical ways. The company guides its new hires through a virtual environment in which they attend an orientation session, sign-up for their employee benefit programs, and learn how to do various tasks, including completing expense reports. The company has multiple islands in Second Life, and routinely holds in-world events and meetings with employees from across the organization and around the world (from small group work sessions to large, auditorium style gatherings) [32].

The TSA (Transportation Security Administration) recently announced that it is looking to use video games, or what the agency terms – "games with a purpose" – in order to train its 40,000 security screeners to improve the accuracy of its screening and improve customer service [33]. With the far-flung nature of its workforce, virtual environments would be a natural for training and simulations for the TSA.

Virtual world environments may be especially useful in disaster planning and emergency preparedness training. Indeed, emergency simulations may be a "killer app" of not just Second Life, but all virtual worlds, in that they are cost and time effective. Robert Furberg, a research analyst for the Center for Simulator Technology at RTI International in Research Triangle Park, North Carolina, observed: "A full-scale (emergency response training) exercise takes a lot of advanced preparation and requires daylong drills - it is expensive and time-consuming. With simulation, we can run through a mass casualty event and change the parameters. Each case is a little different, and it is available 24/7" [34]. They also allow for participants to join-in virtual training exercises from wherever they are, saving a great deal on travel costs for such simulations. Virtual simulations allow for scenarios to be practiced and rerun repeatedly with differing scenarios and variables at play. And, as with military simulations, machinma video captures of virtual simulations allow for after-action review capabilities to allow for critiques and debriefings after concluding the exercises. Finally, there is the real need for "refresher" training in the emergency preparedness area. Randy Sickmier is the exercise plans manager for the Emergency Management Training, Analysis and Simulation Center (EMTASC) in Suffolk, Virginia – a non-profit consortium between Old Dominion University and 17 private sector firms formed in 2005. He related the value in using virtual world simulations for emergency management training by stating: "You've got a guy driving in to the EOC (emergency operation center) who hasn't sat in that chair in more than six months. During the day, he's the public works officer for Staunton, Virginia, and all of a sudden he's in charge of some aspect of this emergency response. It's not something he does every day. This is where the simulation can be valuable" [34].

There have been exciting recent developments in this area. The California Department of Health Services, aided by researchers from the University of California-Davis Health System, recently set-up a virtual environment to train staffers in procedures for setting-up emergency clinics in the event of a biological attack. Working in a virtualized representation of the California Exposition and State Fair built in Second Life, a dozen staffers participated remotely in the exercise geared to simulate the administration of antibiotics from the Strategic National Stockpile in response to a mock anthrax attack. The Second Life-based simulation was a replication of a real-life disaster simulation held at the same location two years earlier – one that involved 250 state employees and a thousand patient volunteers. The principal investigator on the $80,000 project, Dr. Peter Yellowlees, who is a Professor of Psychiatry at UC-Davis, explained that: "The aim of the exercise was to see if the state could constantly train people in setting up emergency clinics. One big advantage is that they could do this training 24/7 from wherever they are, and you don't have to recruit patient volunteers" [34].

The U.S. Department of Homeland Security (DHS), working with the Emergency Readiness and Response Research Center (ER3C) at Dartmouth College's Institute for Security Technology Studies, recently created a private island in Second Life, named "Response." The Synthetic Environments for Emergency Response Simulation (SEERS) project was established by DHS to prototype new virtual tools to conduct cost effective disaster response rehearsals for emergency preparedness [35].

At the University of Maryland's Center for Advanced Transportation Technology, researchers are developing a virtual world training exercise – encompassing many different traffic scenarios – from minor accidents to major incidents - for use by emergency responders in the I-95 Corridor Coalition. These simulations can now include hundreds of participants' playing-out their real world response functions in the virtual environment, in scenes such as the one depicted in Figure 2 [36]. Michael Pack, who serves as research director of research for the University of Maryland center, observed: "It wasn't until we started to do elaborate demos that the first responders started to realize the true potential" [37]. Training in the virtual environment can be more cost-effective – and realistic - in ways that staged disasters cannot. For example, in the simulated environment, if a first responder fails to put-on his or her safety vest or reflective jacket when approaching the scene of an accident, their avatar may be hit by a car – a negative reinforcement that could not occur in a real-life training exercise [36]. Captain Henry de Vries of the New York State Police noted that: "This project provides incident responders from all disciplines the opportunity to train together in real time for the purpose of learning the latest best practices in incident scene safety, coordination and quick clearance of highway events" [36].

Fig. 2. Screen Capture from Virtual First Responder Simulation from the University of Maryland's Center for Advanced Transportation Technology

V. Virtual Management Training

Virtual worlds and serious games may prove valuable not just training for specific scenarios and learning narrow roles, but as a way of testing and developing leadership skills that can translate into "real world" managerial jobs [38]. A May 2008 cover story in the *Harvard Business Review* focused on just this subject, examining how experience leading guilds in World of Warcraft and leading other forms of virtual teams in multiplayer online games is now translating into leadership skills that can be transferred to today's organizations. In fact, a study of over one hundred IBM business team leaders who also led groups in MMOGs found this to be the case [39].

The advance of the culture of gaming will require changes in the very way we think about organizations. Leadership in games is "a task, not an identity," and thus, a person does not assume a leadership role with any permanence. In fact, one's role can alternate between leadership and followership. Also, organizational cultures will need to change in a significant way as well – such as a willingness to accept initial failures in order to achieve ultimate group success. However, the speed and pace of game play may be ideally suited for training how to lead people in today's increasingly fast-paced, technologically-intense environment [39].

MMOGs help foster and develop what has been deemed as "the gamer disposition" [40]. And today, this gamer mentality is something that we need to foster - and not just allow, but encourage - both in the educational process and in organizations. The gamer disposition has been described as having five key attributes:

1. Being bottom-line oriented
2. Understanding the power of diversity
3. Thriving on change
4. Seeing learning as fun
5. Marinating on the "edge" (seeking new ways of solving problems).

Employees with the gamer disposition will be "flexible, resourceful, improvisational, eager for a quest, believers in meritocracy, and foes of bureaucracy" and as such, they are more able to succeed in today's workplace than their non-gamer counterparts and have "exactly the disposition you should want in your workforce" [40].

VI. Conclusion

It was recently observed that: "Each generation is comfortable with a different set of technologies. Just wait for those that use games to move into positions of power" [22]. The passage of time will see the rise of the gaming generation into positions of prominence. With this generational succession, there will be a growing acceptance of computer simulations and virtual worlds. Thus, the use of serious games and virtual worlds by the next generation of leaders – comprised of digital natives - will come about because they will simply "expect and demand it," as they: "will not, cannot, manage information on paper, or in spreadsheets or online dashboards. They will not endure the kludgy, slow, inefficient process of learning new software and keeping that knowledge up to date merely to be able to manipulate data….They will demand to see and touch and manipulate what is known about problems and to 'play' possible solutions so they can view the likely outcomes before choosing how to proceed" [41]. Thus, as the gamers come to the fore and the new, tech-savvy generation rises to positions of prominence, it will be imperative for organizations – and colleges and universities as well – to challenge the way things are done to adapt to the needs of the digital native generation.

References

[1] DFC Intelligence (2007), *Press Release: Worldwide Market Forecasts for the Video Game and Interactive Entertainment Industry*, September 18, 2007. Retrieved March 2, 2008, from http://www.dfcint.com/wp/?p=189.

[2] P. Verna, "Video gaming audience is fragmented," *eMarketer*, April 4, 2008. Retrieved August 7, 2008, from http://www.emarketer.com/Articles/Print.aspx?id=1006147&src=print_article_blue_bottom_article&xsrc=print2_articlex.

[3] Mediamark Research & Intelligence, *Press Release: Gaming Nearly Ubiquitous among Kids Online, One-third Have Email Address*, December 21, 2007. Retrieved January 2, 2008, from http://www.marketingcharts.com/direct/gaming-nearly-ubiquitous-among-kids-online-one-third-have-email-address-2839/.

[4] W.J. Au, "All politics is virtual," *The Huffington Post*, March 12, 2008. Retrieved August 13, 2008, from http://www.huffingtonpost.com/wagner-james-au/all-politics-is-virtual_b_91139.html.

[5] D.A. Williamson, "Kids, teens and virtual worlds: It's a virtual, virtual world after all," *eMarketer*, September 25, 2007. Retrieved September 26, 2007, from http://www.emarketer.com/Article.aspx?id=1005410&src=article1_newsltr.

[6] L. Rawlinson, "Virtual worlds: The next Facebook?" *CNN.com*, August 8, 2007. Retrieved October 1, 2007, from http://edition.cnn.com/2007/TECH/08/07/virtual.living/index.html.

[7] J.V. Last, "Get a (Second) Life!: The avatars are coming," *The Weekly Standard*, October 1, 2007. Retrieved October 16, 2007, from http://weeklystandard.com/Content/Public/Articles/000/000/014/145mliuh.asp.

[8] C. Mims, "Second Life chairman's stump speech takes us down the

rabbit hole," *Scientific American Observations*, June 29, 2007. Retrieved August 15, 2007, from http://blog.sciam.com/index.php?title=second_life_chairman_s_stump_speech_take&more=1&c=1&tb=1&pb=1.

[9] L. Gartner, *Press Release: Gartner Says 80 Percent of Active Internet Users Will Have a "Second Life" in the Virtual World by the End of 2011*, April 24, 2007. Retrieved September 13, 2007, from http://www.gartner.com/it/page.jsp?id=503861.

[10] N. Wilson, "Simplicity and the virtual tipping point," *Metaversed*, November 6, 2007. Retrieved September 19, 2008, from http://metaversed.com/tags/news/virtual-worlds.

[11] E. Steel, "Marketers explore new virtual worlds: Some create own as Second Life site loses some luster," *The Wall Street Journal*, October 23, 2007. Retrieved October 30, 2007, from http://online.wsj.com/public/article/SB119309412255567685.html?mod=dist_smartbrief.

[12] L. Lorek, "Kids flock to sites especially designed for them," *San Antonio Express-News*, July 2, 2007. Retrieved August 19, 2008, from http://www.mysanantonio.com/business/stories/MYSA070107.1R.VW4Kidsmain.2a76afe.html.

[13] B. Barnes, "Web playgrounds of the very young," *The New York Times*, December 31, 2007. Retrieved August 3, 2008, from http://www.nytimes.com/2007/12/31/business/31virtual.html?_r=2&oref=slogin&oref=slogin.

[14] R. Holmes, "Where fantasy and reality collide," *The Daily News-Tribune*, July 1, 2007, 2007. Retrieved August 29, 2007, from http://www.dailynewstribune.com/opinion/x360439247.

[15] K. Dell, "How Second Life affects real life," *Time*, May 12, 2008. Retrieved June 14, 2008, from http://www.time.com/time/health/article/0,8599,1739601,00.html.

[16] T. Weiss, "Leading your life in second worlds." *Trendsspotting*, October 29, 2007. Retrieved January 27, 2008, from http://www.trendsspotting.com/blog/?p=243.

[17] M. Prensky, "Digital natives, digital immigrants," *On the Horizon*, vol. 9, no. 5, September/October 2001, pp.1–2.

[18] I. Jukes and A. Dosaj, "The differences between digital native learners and digital immigrant teachers," *Apple.com*, February 2008. Retrieved March 11, 2008, from http://www.apple.com/au/education/digitalkids/disconnect/landscape.html.

[19] M. Prensky, "Learning in the digital age," *Educational Leadership*, vol. 63, n. 4, December 2005/January 2006, pp. 8-13.

[20] J.P. Gee, "High score education: Games, not school, are teaching kids to think," *Wired*, May 2003. Retrieved April 10, 2008, from http://www.wired.com/wired/archive/11.05/view.html?pg=1.

[21] P. Perla, *The Art of Wargaming*. Annapolis, Maryland: Naval Institute Press, 1990.

[22] R. Smith, *The Long History of Gaming for Military Training*. U.S. Army Simulation, Training and Instrumentation Program Executive Office, March 2008. Retrieved April 18, 2008, from http://www.peostri.army.mil/CTO/FILES/RSmith_Military_Game_Training.pdf.

[23] A. Derryberry, *Serious games: Online Games for Learning – An Adobe White Paper*. February 2008. Retrieved February 21, 2008, from http://www.adobe.com/resources/elearning/pdfs/serious_games_wp.pdf.

[24] A. Greenberg, "Building the feds' virtual war room," *Forbes.com*, March 20, 2008. Retrieved June 27, 2008, from http://www.forbes.com/technology/2008/03/20/cyberspies-virtual-feds-tech-security-cx_ag_0320virtual.html.

[25] C. Calongne and J. Hiles, *Blended Realities: A Virtual Tour of Education in Second Life*, November 2007. Retrieved May 10, 2008, from http://edumuve.com/blended/BlendedRealitiesCalongneHiles.pdf.

[26] Anonymous, U.S. Army creates video game squad, *GamePolitics.com*, December 12, 2007. Retrieved December 18, 2007, from http://gamepolitics.com/2007/12/12/us-army-creates-video-game-squad/.

[27] L. Graves, "A Second Life for higher ed: A virtual world offers new opportunities for teaching," *U.S. News & World Report*, January 10, 2008. Retrieved January 18, 2008, from http://www.usnews.com/articles/education/e-learning/2008/01/10/a-second-life-for-higher-ed_print.htm.

[28] M. Wagner, "Turning work into play is no game," *Information Week*, February 22, 2008. Retrieved August 26, 2008, from http://www.informationweek.com/blog/main/archives/2008/02/turning_work_in.html.

[29] E. Castelli, "It's all how you play the game," *Federal Times*, January 28, 2008. Retrieved January 30, 2008, from http://federaltimes.com/index.php?S=3333068.

[30] K. Graft, "US Army founds gaming office," *Next Generation*, December 13, 2007. Retrieved January 30, 2008, from http://www.next-gen.biz/index.php?option=com_content&task=view&id=8337&Itemid=2.

[31] M. Peck, "U.S. Army embraces games - sort of," *Training & Simulation Journal*, December 24, 2007. Retrieved February 1, 2008, from http://www.tsjonline.com/story.php?F=3115940.

[32] D. Kirkpatrick, "It's not a game: The 3-D online experience Second Life is a hit with users. IBM'S Sam Palmisano and other tech leaders think it could be a gold mine," *Fortune*, February 5, 2007, vol. 155, no. 2, pp. 56-62.

[33] W. Chan, "Learning could be fun and games at TSA," *Federal Computer Week*, July 31, 2007. Retrieved January 18, 2008, from http://www.fcw.com/online/news/103369-1.html.

[34] D. Raths, "Virtual worlds help public safety officials practice for real-life threats," *Government Technology*, March 27, 2008. Retrieved April 2, 2008, from http://www.govtech.com/gt/261426?topic=117693.

[35] J.R. Aitoro, "Uncle Sam gets a (Second) Life," *CMP Channel*, May 8, 2007. Retrieved August 26, 2007, from http://www.crn.com/showArticle.jhtml?articleId=199400068.

[36] The Center for Advanced Transportation Technology (CATT), *Press Release: Preparing for Real-life Disasters in the Virtual World: Clark School Develops Online Training Environment for Emergency Responders*. University of Maryland, College Park, A. James Clark School of Engineering, October 16, 2007. Retrieved February 18, 2008, from http://www.catt.umd.edu/news/news_story.php?id=2596.

[37] C.G. Lynch, "Companies explore virtual worlds as collaboration tools," *CIO*, February 6, 2008. Retrieved March 20, 2008, from http://www.cio.com/article/print/180301.

[38] C. Gaylord, "Can web-based worlds teach us about the real one?: Researchers are eager to study millions of online gamers that live in 'virtual' communities such as Second Life, hoping to gain insight into our 'real world' behaviors," The Christian Science Monitor, January 23, 2008. Retrieved April 11, 2008, from http://www.csmonitor.com/2008/0123/p13s01-stct.htm.

[39] B. Reeves, T.W. Malone, and T. O'Driscoll, "Leadership's online labs," *Harvard Business Review*, vol. 86, no. 5, May 2008, pp. 58-66.

[40] J.S. Brown and D. Thomas, "Breakthough ideas for 2008: The gamer disposition," *Harvard Business Review*, February 2008. Retrieved January 30, 2008, from http://harvardbusinessonline.hbsp.harvard.edu/hbsp/hbr/articles/article.jsp?articleID=R0802A&ml_action=get-article&print=true&ml_issueid=BR0802.

[41] A. Laurent, "Be the data," *Government Executive*, January 25, 2008. Retrieved February 3, 2008, from http://blogs.govexec.com/techinsider/archives/2008/01/be_the_data.php.

Active-based Key-skills Learning in Engineering Curriculum to Improve Student Engagement

Cecilia K.Y. Chan
The University of Hong Kong
Centre for the Advancement of University Teaching
Hong Kong

Mike Murphy
Dublin Institute of Technology
Faculty of Engineering
Ireland

***Abstract* – Employer groups have been extensively lobbied the government to introduce key skills development strategy for higher education to produce well-rounded graduates. This is particularly true in professional disciplines such as engineering where discipline specific knowledge rapidly becomes obsolete. Key-skills is the key to success; they help one to excel in the workplace and their importance cannot be denied in this information age. However, it is difficult to teach and assess key-skills as students are not familiar with the importance and relevance of key-skills and lecturers are not familiar how to embed key-skills into the traditional core discipline teaching. In this paper, we will discuss the changes in an electrical engineering programme in Ireland - how curriculum reform has given an opportunity to embed key-skills into the programme. We will also discuss the research investigated on how to engage students to key-skills learning and several development strategies used to demonstrate key-skills in an engineering curriculum.**

I. Introduction

The Irish Government is constantly examining the Irish Education system to ensure a continual supply of qualified graduates with Science, Engineering and Technology (SET) skills, since if current trends in the supply of skills by the broad education and training sector are left unchecked, there will be a significant short-fall in skills required to fill this growth. From the report [1], it was found that job skills in sectors which depend heavily on science and technology will be the most crucial in order to ensure that Ireland is well-placed to benefit fully from the upturn and to be at the vanguard of the global move towards a knowledge-based economy. The basis for future growth in prosperity and jobs is by further investment in the knowledge, skills and innovation capacity up to fourth level that will drive economic and social development in an increasing competitive global environment. In fact, this is true not just for Ireland but for most developed countries. The higher education system must deliver people who will expand knowledge-based business in Ireland [2]. In addition, the Irish government must constantly examine the required skills so to reflect the changing needs of the enterprise sector.

It is a fact that Ireland's economic prosperity future will primarily depend on the SET sectors. However, from the College Application Office statistics (CAO Statistics 2000 – 2006), there has been a 41% drop in the students' 1st preferences in the Engineering/Technology course group in fig. 1.1 from the period 2000 to 2006, and although the Science/Applied Science course group in fig. 1.2 has remained stable but there has been little increase in the number of students choosing Science as their 1st preferences.

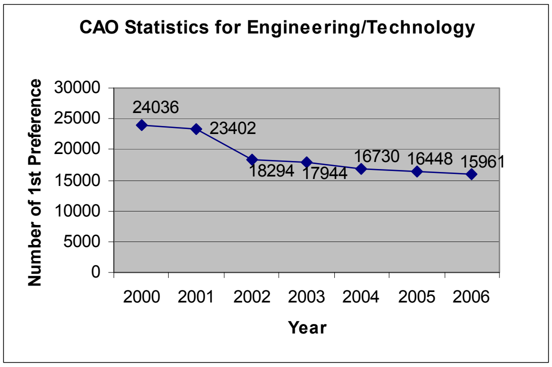


Fig. 1. Number of 1st Preferences for Engineering/Technology in CAO

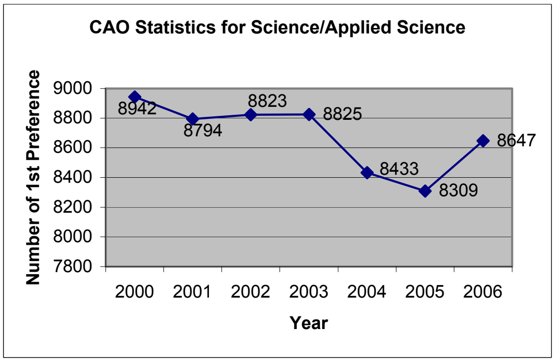


Fig. 2. Number of 1st Preferences for Science/Applied Science in CAO

This is despite the endless number of educational and government reports highlighting the significant skill shortage in the SET sectors; the large amount of funding and scholarships introduced by the government to attract more students into the SET areas from primary degree through to PhD level.

Apart from the external forces to recruit and retain more students into SET, there is also an external force from the employers to introduce key-skills development strategy for higher education. Key-skills or soft skills, often referred as transferable skills can be defined as the basic skills which *'give*

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_16, © Springer Science+Business Media B.V. 2010

an individual a secure foundation for life and work. They thus cover vocational and technical skills, as well as social and personal competencies' [3]. Most of the business leaders commented that they could find employees who have "hard skills", that is the capability to operate machinery or to fulfil other tasks, but many potential hires lack the "soft skills" that a company needs. Employers are now looking for well-rounded graduates who do not just excel in their disciplines but on a variety of broader subjects and skills such as communication, teamwork, initiatives, problem-solving, self and time management and lifelong learning; "we are increasingly articulate about what we expect from our university students. We expect graduates to be able to manipulate subject knowledge critically and analytically, argue effectively and reflectively, organise their study and time, work well in groups and so on" [4]. CEOs and human resource managers are also favouring employees who can demonstrate a high level of keyskills over the employees who have a high level of technological knowledge. They believe hard skills can be trained easier than soft skills, and in a world where discipline specific knowledge rapidly becomes obsolete, it means employees need to upgrade their hard skills regularly as hard skills would only have a short shelf live, thus employees with good lifelong learning skills and positive attitudes towards learning are preferred.

Soft skills play a vital role for professional success [5]; they help one to excel in the workplace and their importance cannot be denied in this information age. After all, as Peter Salovey the psychologist who invented the term E.Q. explained "*I.Q. gets you hired, but it is E.Q. get you promoted.*" Good soft skills (which are in fact rather scarce in this highly competitive world) will help the graduates to stand out in a milieu of routine job seekers with mediocre skills and talent. In fact, if students are to be successful, we must refocus the curriculum to prepare students to third level learning, to help them acquire the skills to navigate a complex organisation, and in so doing, to embrace processes of learning, personal development and career planning for life at university and beyond graduation.

II. Difficulties with Learning and Teaching Key-Skills

Key-skills cannot be developed or learnt merely by reading textbooks. They must be shown, demonstrated, explained, made aware and given an opportunity to demonstrate and practice [6]. It is obvious to us – Educators, why transferable skills have been high on the academic agenda but for the green students, the reason may not be so apparent. Most engineering students do not realized the importance of key-skills, they do not see the relevance of it. As you could imagine, seventeen or eighteen years old students who choose to become engineers are usually because they like the idea of designing, building and creating. Their goals are to become successful engineers. Students often feel these non-discipline subjects which are totally unrelated to their main discipline are taking up their time and efforts away from their main goals. It is important to remember engineering itself is not an easy discipline and the contact hours each week in average is at least 30 hours in any higher education institutions. Thus some have found the lectures on professional development or communication studies revolting and uninteresting, many have even commented that, "*I didn't come here to learn how to talk and write.*" To summarize – they detest it and this detestation may drive them to withdrawal.

In addition, key-skills is not by tradition part of the engineering curriculum. Service courses such as professional development or key-skills courses are often taught by an outside faculty due to university policy. The home faculty has little choices on the type of lecturers who take on these courses. Many academic professors do not actually have the personality or skill-sets to demonstrate such wisdom. A communication course might be taught by an elderly professor whose expertises are English grammars and essays, with little practical knowledge on skills such as presentations, team-working, leadership communications and IT skills. They are not prepared to upgrade their understanding of the word communications, emailing and other high-tech related communication techniques which are now deemed as gadgets of any genuine professionals. Often the learning methods are not relevant to the skill-sets they try to teach or examples are too outdated and not of an authentic nature. And of course, it doesn't help when the students are not properly explained what the learning outcomes of such skills will delivered.

The question that we asked ourselves is how can we transfer our belief to them that key-skills is the answer to employability and how can we make it more interesting for these youngsters.

III. Relating the Importance of Key-Skills to Students

To explain the importance of key-skills, we have to teach and associate key-skills by relating what is importance to the students. But what is important to them? In order to answer that we must think like them and the question "why do we want to go to college?" is likely the first question raise in many students' minds prior to apply and enter college. It is a very important question that has many different answers for many different people. The common reasons for attending college are usually within the followings:

1. Better career paths – this is probably the most practical and realistic reasons why most students go to college. There have been many statistics showing a college graduate tends to make more money than a non-college graduate. And better job opportunities tend to lead to better life styles.

2. College experiences – the college experience allows a person to broaden their horizon and to grow as a person. There is not many place in the world, you can learn such a unique experience, a place where you can redefine yourself, meet people with different backgrounds and learn esteem for oneself and for others. Obviously, students don't usually think that deep prior to entry and relate their college experience with the whole idea of the developmental cycle. They tend to anticipate the college experiences in relation to freedom, independence and socialization with peers – a step closer to adulthood.

3. The norm – many students go to college because it is the norm these days, for them, from kindergarten to primary school to secondary school and to university is just the next step of life.
4. To please others – often individuals go to college to please parents or love ones. Unfortunately, pleasing others is often not the best reason to go to college, without self-commitment, it is likely to be short-termed.
5. Knowledge – to learn. Unfortunately, most students who go to college are not primarily going there to learn, this type of intrinsic motivation occurs more in mature students.

Majority of the students who go to college are because of the combination of some of the above reasons but the most common reason is ultimately to gain a good career. This is evidenced by a student information gathering sheet which was surveyed on the first day of the first years from two full-time electrical engineering programmes at the Dublin Institute of Technology (DIT), Ireland (the author was an engineering assistant professor and a programme coordinator at DIT prior to taking up her employment at the Centre for the Advancement of University Teaching at the University of Hong Kong). The information gathered includes information such as email address, phone no, home and term address, hobbies, IT skills and one open-ended question "In one sentence, describe what you want to achieve in life?" Data were collected from more than 250 students for 5 years from the year 2003 to 2008. These students aged from 17 years old to 40 years old with diverse backgrounds, about 70% of the students surveyed came directly from secondary schools and are first-time university entrants. Out of the 250 students, more than 70% of the valid data shows a similar answer "I want to be successful and happy", sometimes the words "rich" and "have a family" will also be included. This result indicates career opportunity is the ultimate reason why a student goes to college. [We should note career opportunity is not necessary the most important reason for student retention because students sometimes fail to remain such long term commitments.]

IV. Curriculum Reform - An Opportunity to Redesign with Keyskills

The programme in question was a 3-years full-time Electrical Engineering Technology programme at DIT. Some basic key-skills in the areas of writing, oral presentation, work group and organisational communication and IT skills were taught in first and second year and project management was taught in third year; these courses are compulsory core courses. Unfortunately, there has been serious difficulty engaging the students in these courses particularly in the communication area of the course. Data was recorded for 5 years in a roll and was shown that the attendances – an average class of 40 students dropped from full attendance in the first two weeks to less than 15% of the class attending by the end of the first semester. Interviews with students had indicated that they are not interested in the courses and found them "unrelated and boring". This clearly echoed the difficulties of learning and teaching key-skills we discussed earlier.

Some suggestions were made by the lecturers involved to improve these courses to engage student learning. However, most professors are familiar that any changes in a curriculum of any course would require formal consent from the programme committee and likely from the faculty board. The process can be lengthy and cumbersome, and unless a valid reason is put forward for the changes, it may not necessary be granted, thus academics tend to avoid any changes necessary.

With the renewal of the accreditation from the professional body, it has given us a window of opportunity for a curriculum reform; to re-evaluate and renew the curriculum. In fact, the engineering professional body [8] embraces many key-skills criteria in their expected programme learning outcomes, namely:

1. **Discipline Knowledge** – Knowledge and understanding of the mathematics, sciences, engineering technology sciences and technologies related to the particular branch of engineering technology.
2. **Problem solving** – Ability to identify, formulate and solve broadly - defined problems in the particular branch of engineering technology.
3. **Design and create** – Ability to contribute to the design of components, systems and processes to meet specified needs.
4. **Research** – Ability to conduct investigations to facilitate the solution of broadly-defined problems within the particular branch of engineering technology.
5. **Ethics** – Understanding of the need for high ethical standards in the practice of engineering, including the responsibilities of the engineering profession towards people and the environment.
6. **Teamwork and Lifelong Learning** – Ability to work effectively as an individual, in teams and in multi-disciplinary settings together with the capacity to undertake lifelong learning.
7. **Communications** – Ability to communicate effectively with the engineering community and with society at large.

From the list above, apart from criterion one which is specifically related to the engineering discipline, the other criteria are all in some way related to key-skills. This has given us some indications for re-development in the topic of key-skills.

V. Curriculum Reform with Keyskills

With the opportunity to examine the curriculum and the understanding that students realized good career prospects as their ultimate goal, we began to investigate the learning outcomes specified by the professional body to align with appropriate learning activities and assessments. In line with modern trends in education delivery when developing the programme, significant modifications have been made to the assessment methodologies applied with emphasis on end-of-year examinations reduced and greater emphasis being placed on continuous assessment elements throughout the programme. It was agreed by the programme committee that the professional development and project management courses would be 100% formative assessed, which would also allowed more flexibility on the delivery of the courses.

A. Embedding Key-skills into Mainstream Curriculum

Researches [9] have shown that by integrating the development of soft-skills into the mainstream curriculum would provided more relevance and variety to the mainstream core subjects (less overwhelmed by the theoretical context) and would also allowed these skills to be practiced and demonstrated more pragmatically.

Most students choose to become engineers because they like the idea of inventing, designing, building and creating "their innovative visions". But sometimes, their early experiences in college can lead to expectations which are quite the opposite, as there are lots of theoretical studies and set practical "follow the procedures" labs involved, often students would not encounter any real design work until final year, there are scarcely any wild, brainstorming sessions in the duration of their early years. For some students, these experiences can be quite a turn-off. They forget the reasons why they like engineering and moreover, begin to question their choice of career. Some people say college can make us less intelligent. Is that true or does college cause us to lose parts of the creativity in us because we become too adapted to think in a particular way? Students can become too accustomed to the way they are being taught - learning by rules and by theories, and forget the real him or her.

Traditional labs are obligatory for students to learn and comprehend the fundamentals but at the same time, students should be encouraged with their creative ideas and never disheartened even if they are wrong, because it is alright to make mistakes. After all, Rome was not built in one day. Only with trials and errors can new inventions be invented.

Thus, it was decided that soft-skills such as creativity, self-directness and critical thinking should be introduced in their early years of studies. In the first year of the programme, a design project competition was introduced to explore their natural creativity in a freeform, brainstorming manner rather than a structured formal approach. The project mimicked the television show "Scrap-yard Challenge". Students were to implement and build an engineering device which related to a real life application by trial and error, given a common theme and a set of rules. Students had an option to work in pairs or by themselves. They were giving six weeks to complete the project and two 2-hours workshops were setup to assist with their ideas and construction. The use of recycling materials was also greatly encouraged. Initially, it was agreed that the project design should not form part of the assessment of the course and be optional to reduce any pressure and alteration of the degree programme. And even though 80% of the class did participate due to the competitive nature and also their interests in developing their own ideas, it was later decided to implement it as part of the course curriculum, and made it mainstream as it would be beneficial for all students. The assessment included a presentation on the design process and a demonstration on the end-product. The winner was assessed on the overall design, originality, feasibility, the economical and environmental aspects and the presentation skills; these criteria were explained and given to the students [10].

The project was competitive as the award was very inspirational, it was a summer job contract with a major electrical engineering company, the winner was judged by the professors and a representative from the company and must satisfied the rules and conditions of the project. Winners and runner-ups were invited to university open days and engineering fairs to demonstrate their enthusiasm in the subject and ultimately, attract and introduce new applicants to the field.

The result was encouraging as some of the end-products were surprisingly innovative even by the professors' standards; in fact, both students and lecturers were motivated by the results. A quote from one of the professors: "I can't believe they build that, it is amazing and it works!" The project competition has been running for 3 years now, and observations and discussions with the students had clearly shown that the students are indeed more engaged at this type of design project. During the project, certain key-skills were made aware and given opportunities to practice and demonstrate towards the goals they believed are important – their career opportunities.

Following the success of the design project, the professional development course in second year was replaced by a design course which involves in designing a robot using a DSP chip. This course is similar to the first year design project except workshops are held every week for one semester to give students some basic backgrounds including how to programme the DSP chip. Students are expected to work in teams in this design project, which added an extra level of key-skills. Each team was to hold a journal logbook to report on weekly progress; any members of the team can enter data and comments in the logbook.

Embedding key-skills learning into core curriculum has significantly helped student engagement and widening student participation both on professional development courses and core engineering courses. More importantly, it has allowed them to use their creativeness, self-directness and self-motivation to establish balances between their expectations as engineers and the less-exciting learning facts in college; maintaining student interests both as a college student and as a role of an engineer to assist student motivation and retention.

B. Career Planning and Lifelong Learning as Key-skills

Career Management Skills are '*the set of skills that required in order to maximise your employability in an organisation; knowing your strengths, interests, values, and skills and able to act upon where areas are weak to maximise your employability and widen your opportunities*.'

Career planning does not seem to be a word that occurred greatly throughout education, any career related courses are usually organised by the career services in the university which are optional to attend and not discipline specific. There is little recollection on serious thoughts or planning revolving the word 'career'. In fact, luck, faith and circumstances following the common steps in life may be the words that played larger parts of many people's career and life to date. This may be possible in the past era but not for the generation present and future, as the work ethic is changing. Graduates no longer have one job for life, often graduates become self-employed once they graduated; they need to plan for changes and cope with uncertainty and more importantly, they need to continue to learn to stay employable. Students need to know their path socially, personally and professionally; to achieve that they need career planning and lifelong learning skills.

The Career Services in DIT was invited to set up a one-day career planning workshop for the senior years of the degree students in Electrical Engineering. It was decided that the Career Services was best to take on the workshop role as they have experts in the area and would able to integrate the engineering aspects into the career education. The learning outcomes of the workshop are to

- Reach a competent level of self-awareness in relation to their skills, aptitudes, personality and interests.
- Develop knowledge of decision-making techniques and apply them in creating a personal career plan.
- Complete research into the engineering occupational profiles and the range of opportunities available.
- Promote themselves effectively in written and verbal communications, e.g CVs, application forms, interviews and presentations.
- Network effectively and maximise exposure to career opportunities both advertised and speculative.
- Develop the ability to critically evaluate personal progress and adapt a strategic approach to handling changing circumstances.

The workshop was first organised in February 2005 and was reorganised in early March 2006 and 2007 to extend to more engineering students in the school to learn about key-skills and career directions. The students were surprisingly enthusiastic and motivated by the event and found that the CV writing and interview guide to be exceptionally invaluable. Some of the content in which they were delivering on the day was in fact similar to their normal professional development module. Attendance data were collected and observations were made that students tend to be more engaged with compulsory stand-alone workshop. It differed in the delivery methods which were active learning and that it was held as a special one-day workshop unlike normal lecturing hours. Also it was directly related to their careers. All students who attended the career workshop received certificates and credit. The difficulty was to try to find a suitable time in the timetable to hold the workshop as it was a full-day workshop, ideally we would append this workshop as part of the professional development module.

The programme team hopes to redesign the professional development course module in first year into five full days stand-alone courses which spread across the entire academic year. For example: one full day of professional development course each month. The day will be activity-based using problem based learning, focus group, interviews many-to-one or one-to-one basis, presentation, debates etc as part of the delivery method. It is planned to involve other lecturers as guest lecturers for learning activities and assessment.

C. Demonstrate the Relevance and Significance in Key-skills

The career planning workshop and the curriculum renewal with embedded key-skills had demonstrated the importance of key-skills to students. But as mentioned key-skills needs to be shown, demonstrated and moreover, given an opportunity to practise. In order to have significant impact at the practices, two events were organised to give students an opportunity to demonstrate their key-skills. One event was organised for the junior years and the other event for more senior years. Namely, the Graduate information night and the Career Speed Dating Luncheon.

1. Graduates Career Information Evening

Engineering alumni were invited back to share their experiences when they were in college and now in the real world at an evening event. The setup of the event included presentations by each alumni, this was then followed by an informal social gathering where drinks and food were served. The graduates were first briefed to ensure they don't just focus on the good side of their lives, they would emphasise the difficulties – the 'failure' they may have encountered while they were in college and at work. This is important as students want to know the real side of the story; the most successful story is not when someone did well in college and lived happily ever after; the most successful and interesting story is when a student worked hard, played hard and then succeeded. Sufficient time was allowed with the graduates for question and answer sessions and networking. This event has resulted in huge encouragement booster for students, we should remember graduates maybe just past students to us, but for our students, they are potential employers and contacts, we are giving them an opportunity to demonstrate how to communicate to prospective employers in their field. Three career information evenings were organised in the past three years, some students have continued to network with these contacts and there were about 10% of the students have resulted in receiving summer job offers due to these events. We also gathered data from successful alumni to explore dynamics and experiences that gave rise to effective course completion. From experiences the event would be most beneficial when it occurred just after the first semester when some students are having uncertainties on their chosen field.

2. Career Speed Dating Lunch

The second event – Career Speed Dating Luncheon was targeted at more senior years. It was modelled from the popular speed dating event. This unique event provided students and potential employers with an opportunity to learn more about each other by meeting in a pleasant environment. Organised by the programme chair, students met and engaged with potential employers in a formal business like lunch. This provided a perfect opportunity to allow the students to demonstrate the key-skills that they learnt on an occasion that mattered to them. Following the lunch, a "Speed Dating" setting was adopted for which the students had an opportunity to learn more about the companies and had a one-to-one session in 10-minute intervals. The idea of "*speed dating*" gave a fresh and "*in*" feel to both young students and companies alike. Through this 'speed dating' format, students could afford the opportunity to meet directly with employers, discussed career opportunities, and established a network of contacts in the industry. The event was also of benefit to employers. Employers gained a captive audience. They had the chance to meet one-to-one potential employees and received relevant CVs for positions. They then had a cohort of students with the knowledge and skills they required for positions in the company.

The first pilot "*Career Speed Dating lunch*" was organised in April 2006. Both companies and students gave positive remarks and commented how successful, and well-organised it was. A number of students got summer and graduate job opportunities from the event. The Career Dating event was again organised in March 2007 and 2008. It has attracted many news reporters from the media, newspaper and engineering journals, and career counsellors to attend. And very positive reports were published to the public.

This type of event will ultimately widen student participation and maintain student interest. It helps students to see their future career directions while gives them an opportunity to demonstrate their transferable skills.

VI. Conclusions

Employers and indeed the public these days expect much more from the graduates than just being proficient in the field of studies they engaged in. Life, social and key skills are essential for any employment and indeed promotion. These skills cannot be easily taught and studied in the tradition way, they can only be learned by watching, by realising and demonstrating and then by practising. We should keep in mind that teachers are role-models to students since when we are young we follow what our parent do, and when we are older, we follow what our teacher do. This is particularly true for personal academic tutor who plays a key role in relation to students' soft-skills development.

The learning methodologies we discussed in this paper are practical and active-based. It is believed that by actively engaging the student throughout and rewarding effort accordingly, the learning experience is greatly enhanced and the student is more likely to succeed in the programme and in the real world. It is also important for educators to remember students are not going to find any courses interesting unless we make them interesting, unless we ourselves shown enthusiasm in them.

A throughput of students is important if we are to be a knowledge-based economy, but a throughput of quality and contented students is more important if we are to attract students to engage in 4^{th} Level Education or other Continuing Professional Development beyond their degree graduation.

VII. References

[1] Expert Group on Future Skills Needs,. "Fourth Report of the Expert Group on Future Skills Needs," *EGFSN*, 2003

[2] Fianna Fáil, "Manifesto – Now, the Next Steps," *Manifesto*, 2007.

[3] Council of the European Union "Education Council to the European Council on the Concrete Future Objectives of Education and Training Systems," *Report,* 2001 http://register.consilium.eu.int/pdf/en/01/st05/05980en1.pdf (accessed July 2006)

[4] Staddon. L, "When the new meets the old; the need to re-conceive higher education", *Exchange Journal, HEA,* 2002, pp.28

[5] National Academy of Engineering, "The engineer of 2020: Visions of engineering in the new century," *The National Academies Press,* Washington DC, 2004

[6] Pellegrino, J.W, "Rethinking and redesigning curriculum, instruction and assessment: What contemporary research and theory suggest." National Center on Education and the Economy for the New Commission on the Skills of the American Workforce, 2006 http://www.skillscommission.org/pdf/commissioned_papers/Rethinking%20and%20Redesigning.pdf (accessed June, 2008)

[7] Engineers Ireland, Accreditation Criteria for Engineering Education Programmes, *Engineers Ireland*, March 2007

[8] Chan. C, "Changing the saying -*"Engineers work better with machines than people"* by enhancing key-skills with engineering in the most relevant ways", DIT Learning and Teaching Showcase, 2005

[9] Chan, C. and Coyle, E, "Natural, Creative, Brainstorming "Scrapyard Challenge" Design Project", *Proceedings of 33 International Symposium IGIP/IEEE/ASEE*, Fribourg, Switzerland, Oct 2004.

Assessment & Efficacy of an M-Learning Course in Industrial Hygiene and Occupational Safety

T. J. Ryan
Ohio University
E344 Grover Center
Athens, OH 45701 USA

***Abstract*-This project involved the creation of 20 introductory industrial hygiene and safety video modules specifically for an M-Learning environment. Testing was performed on 9 modules to determine whether this M-Learning content and presentation was as educationally effective as a traditional lecture presentation. Comparison of test scores and key concepts between two groups of college sophomores and juniors was made. The lecture group (n=19) received all course material through classroom presentations, while the video group (n=18) was provided similar content via .mp4 modules delivered via the web (Blackboard). Topics covered included history, legalities, toxicology, airborne hazards, exposure limits, calibration, and sampling methods. There was no significant difference ($p<0.05$) in mean test scores between the two groups (margin of error = +/- 4%). Analysis of responses to a subset of designated essential concept test questions indicated the lecture group performed higher on these topics. Results showed that given the choice, students had a preference for lecture over the .mp4 format. It is concluded that M-Learning, as designed and presented in this course, can provide an effective alternative to traditional classroom lectures in the applied sciences of industrial hygiene and safety. However, instructor emphasis can greatly influence the learning of key concepts.**

I. INTRODUCTION

Technological advances in personal electronics have created distinctive venues for basic and continuing education in industrial hygiene and safety. Pocket sized audio-video players, PDAs, and cellular phones are now available with sufficient memory to hold an entire course of technical content, and students are increasingly utilizing these devices in an evolving form of mobile distance learning dubbed "M-Learning." Reference [1] has described M-Learning as "…any activity that allows individuals to be more productive when consuming, interacting with, or creating information, mediated through a compact digital portable device that the individual carries on a regular basis, has reliable connectivity, and fits in a pocket or purse." The term portable device can include phones, PDAs, iPods or other MP3 players. Reference [2] has reasonably expanded that definition to include "…any sort of learning that happens when the learner is not at a fixed, predetermined location, or learning that happens when the learner takes advantage of learning opportunities offered by mobile technologies."

Although most would agree that modern civilization is poised for a technology-driven explosion of M-Learning opportunities, according to [3], technology enhanced efforts in mobile learning date as far back as the 1970s. In a time that began with cassette tapes the Dynabook was one of the first portable devices, no larger than a notebook, that could store over 500 book pages, or several hours of audio [3]. Ref. [3] also notes the rapid evolution into M-Learning by the turn of the century, with the creation of Laptop PCs and PDAs, which were driving forces for the creation of what we know and use today. Although M-Learning would not be what it is without the advances of the late 1900s, the 21st Century is already proving to surpass all previous technological advances, with the advent of wireless PDAs, 3G service, and even Bluetooth. As history has shown, continued advances in technology can reasonably be expected to improve M-Learning technologies and their ready, affordable accessibility.

Clearly the product of an evolutionary process, M-Learning should now be recognized as a distinct subset of E-Learning. Despite the fact that both terms are often intertwined, the two methods have a sizeable and growing number of differences. In contrast to E-Learning, M-Learning is not only geographically independent, but also environmentally independent, omni-spatial and ubiquitous [4]. Keegan stresses the importance of the mobility factor to distinguish between the two, and goes on to note 4 characteristics of M-Learning technology that can differentiate the two modalities: 1) M-Learning devices are ones that persons typically carry at all times, 2) they regard them as personal devices, 3) they are relatively inexpensive and easy to use, and 4) they can be used almost constantly in most aspects of life and in a variety of different settings [5]. M-Learning can be thought of as a progressive step from E-Learning, in which it has advanced to the next level by focusing primarily on mobility, and those associated factors which enhance its ease of use. It is this mobility that will allow the rapid expansion of M-Learning topical areas outside the domain of the high school or college, and into the area of professional continuing education. As such, content in typical areas such as the applied sciences of occupational health and safety will be increasingly popular.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_17,

While mobility is the hallmark benefit of M-Learning, it also holds the possibility of increasing productivity by making learning available anywhere, anytime, but only as needed [6,7]. According to [8], six added benefits include connectivity, flexibility, interactivity, collaboration, extended opportunities and heightened user motivation [8]. Drawbacks of current M-Learning implementations can include a general lack of expert content, or a dearth of desired topics in the preferred format, as well as network security, a diversity of devices to incorporate, lack of staff expertise, and procurement and maintenance [7]. According to [9] a major reason why M-Learning systems may not have widely proliferated in education to date is because of concerns among faculty and administrators regarding the viability of the mobile devices in online programs. Including devices in the pedagogy raises concerns among faculty as to their usefulness in education [9]. Such questions are entirely legitimate and to be expected. Thus it is necessary for those who would develop M-Learning technology to prove its utility, and demonstrate ways in which serious learning can be accomplished through its use. This is particularly important for adult learning applications, where personal growth and promotion may be directly linked to the completion of certain courses offered in an M-Learning format. Market forces can also be expected to significantly reward meaningful and desired M-Learning content.

The US could hardly be considered a leader in M-Learning deployment at this time, but success stories of M-Learning are beginning to become commonplace. The fields of science, engineering and technology have all experienced success with currently available mobile technology. A case-control report in a computer science course at Georgia Tech demonstrated such positive results. Case students were provided lectures before class on their iPods or laptops, but lecture content, homework, and exams were the same for both case and control groups [10]. The author found that the students who used their iPods or laptops scored about 10 percent better than the group who only attended lectures. A notable benefit of the M-Learning approach was that there was more time for meaningful classroom discussion, since more basic material had previously been conveyed through the M-Learning approach. At Duke University a general physics class has also reported M-Learning styled success. Students were provided video lectures containing audio and images captured from a tablet PC, videos of demonstrations either done in class or available from other sources, and brief introductions to many of the labs [11]. While qualitative in nature this report does, at least, indicate the benefit of M-Learning even to traditional on-campus students who are unable to attend a given class, who may need extra help, or have special needs.

The future of M-Learning presently seems most closely allied with that of the latest generation of cellular phones. According to [12], US developments related to the usage of cell phones portends significant M-Learning increases. Already in the US, the penetration of mobile phones in many junior high schools is 40%, and 75% or higher in many high schools. Where M-Learning is perhaps most likely to be innovated—the college campus—fully 90% of the students utilize the technology [12]. In many parts of the world, cellular usage is in fact much higher and/or advanced. In Japan for example, practically 100% of all students and adults own a cellular phone, and 60% of that technology is already at the 3G level [13]. Cell phone usage is ubiquitous throughout Asia, where Hong Kong and Taiwan have reportedly surpassed 100% penetration levels [12]. Cellular usage is growing significantly in lesser developed countries as well. Kenya and the Democratic Republic of the Congo, for example, have already begun M-Learning applications delivered via cellular networks [14]. In all of these areas, it is becoming clear that advanced cellular phones will be the platform for M-Learning, as applications formerly provided on PDAs migrate to this connected, mobile platform [7]. As implemented the devices already offer a suite of user applications formerly exclusively the domain of the PC or a dedicated unit. Examples include (mobile) banking access [14]; entertainment such as television episodes, movies, and music; games; and of course, communication. Into this mix will come M-Learning, with all of its various approaches and content areas. For the busy professional in the field of occupational safety and health, it will be imperative that the M-Learning they access be expertly prepared, tested, and of the highest quality. To achieve such ends, all M-Learning must be rigorously assessed for accuracy and effectiveness as deployed. This paper reports on actions in that regard to one specific application.

II. METHODS

The study group consisted of a total of 37 sophomores and juniors enrolled in an introductory course in health and safety at the host institution. Students were given the option of not attending class and obtaining all lectures exclusively via the web, for the first three weeks of the quarter (until the first midterm exam). Once selected, those students were instructed not to attend class again until the first midterm exam. Eighteen students self-selected for this Video Lecture (VL) option, while the remaining 19 students formed a control group termed Classroom Lecture (CL).

The control group was slightly better in terms of past academic performance, with an average GPA of 3.1 as compared to the VL group (GPA=2.8). The CL group was comprised of 6 males and 13 females whereas the VL group was almost perfectly gender reversed, with 5 females and 13 males. Both groups had similar past exposure to online learning. A presentation of both groups' characteristics is shown in Fig. 1.

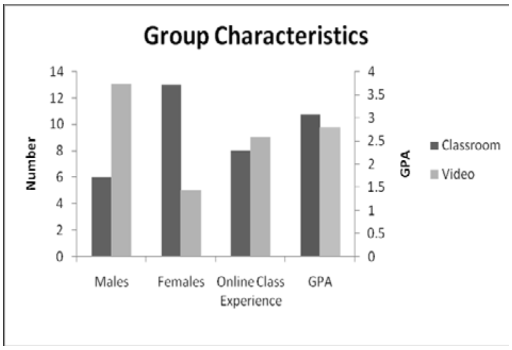


Fig. 1. Classroom Lecture versus Video Lecture Group Characteristics.

The VL group was expected to obtain all course content in the form of nine (9) .mp4 videos and associated PowerPoint slideshows delivered via Blackboard to their student account. Students were able to import the videos onto portable devices or watch them on a computer, although the resolution and screen size of the content was designed exclusively for the smaller format typified by the Apple® iPod. Those CL students who attended each class would receive approximately 6 hours of on-topic classroom contact, as compared to the 2 hour 40 minute run time of the videos for the VL group. No measure was made of the number of times the VL group watched each video.

Three aspects of the M-Learning content were assessed in this study. To determine the educational efficacy of the content presented to the VL group, a comparison was made of the two groups' performance on the first midterm exam. (In the event the M-Learning content proved to be wholly inadequate, this pilot testing was limited to only the first third of the course.) The assessment included a subset evaluation of the groups' performance on 13 "essential concepts", as determined by the instructor. Qualitative aspects of the .mp4 video experience were also examined, via a 9 item questionnaire provided to the VL group. Finally, the CL group was queried as to their reluctance to participate in the VL option

III. RESULTS

Statistical comparison between case-control groups indicated no difference (t-test; $p<0.05$) in performance on the first midterm exam by the two groups. The VL group scored slightly lower than the CL group, with a group average of 69% correct versus 75% for the CL students. The margin of error associated with the test instrument was +/- 4% (percent relative, 95%), or 1 test question. With respect to the essential concepts questions, the VL also underperformed the CL group (75% correct versus 81% for the CL students). Again this overall difference was not significant ($p=0.41$) with a margin of error of +/- 7.6%. For 3 of the 13 essential concepts test questions, the VL group significantly underperformed the CL group ($p<0.04$—$p<0.001$). The contribution of these questions to the overall difference reported was not further assessed however, and so it is not possible to quantitatively determine the role instructor discussion in the CL group, or total viewing time in the VL group, might have played in these outcomes.

Idiosyncratic results as to the benefits of the M-Learning content were seen. The majority of VL users found the .mp4 content of high quality (Fig. 2) and the downloading experience easy and not a factor in their test outcomes (Fig. 3). Surprisingly, the majority of students in this same group felt they would have done better had they attended class lecture in person, and furthermore, that they would prefer not to continue to receive course content exclusively via the .mp4 videos and Blackboard content (Fig. 4). Despite such perceptions, benefits of the VL approach were appreciated by the group. In response to open-ended questions about their experience, most liked the autonomy of the VL option, as well as the ability to rewind or re-watch the videos at will (Fig. 5).

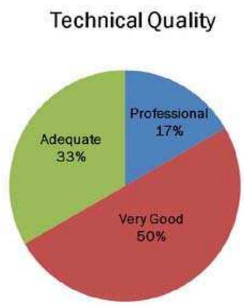


Fig. 2. Video Lecture Group Perceptions of .mp4 Content and Quality.

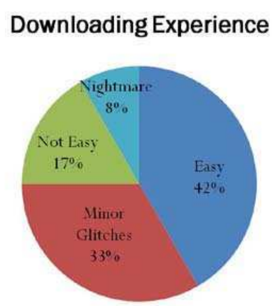


Fig. 3. Video Lecture Group Perceptions of .mp4 Utilization Experience.

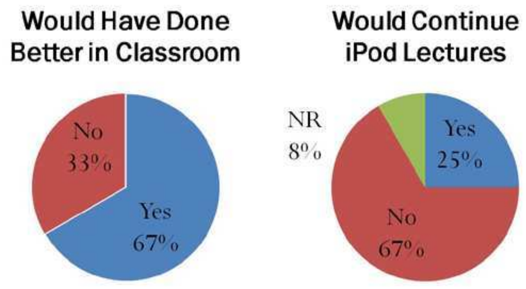


Fig. 4. Video Lecture Group Responses to Perception of Performance Potential (left) and Desire to Continue to Utilize .mp4 Videos Exclusively (right).

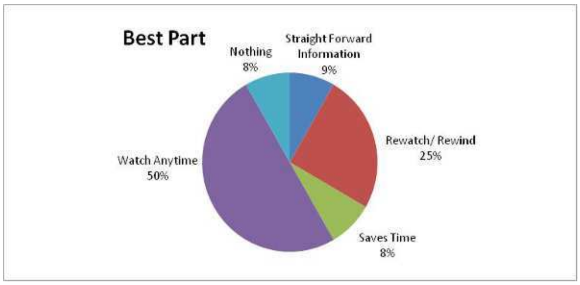


Fig. 5. Summary of Open Ended Responses Concerning the "Best Part" of the Video Lecture option.

Responses by the CL group to the open-ended question concerning their reluctance to participate in the .mp4 trial are summarized in Fig. 6. The most prevalent reason given was that this group of students overall felt more comfortable and better suited to learning in the classroom environment (approximately 56% of respondents). Lesser reasons included previous bad online experiences (5%), lack of equipment (5%), fear of falling behind (17%), and perhaps being distracted by the learning technology (17%).

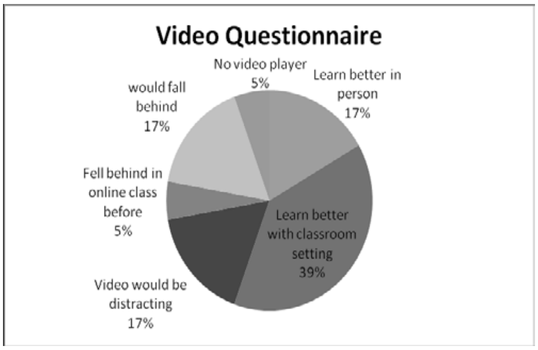


Fig. 6. Summary of Open Ended Responses as to Why the CL Group Did Not Elect the VL Option.

IV. Conclusions

As designed and presented in this course, it is clear that M-Learning can provide an educationally effective alternative to traditional classroom lectures in industrial hygiene and occupational safety. Development of the modules assessed here involved a substantial time commitment on the part of the instructor. Efforts toward added M-Learning content will necessarily need to be directed at high demand rubrics or markets.

This was a very small scale study with some inherent weaknesses. Future work should capture the actual amount of time students in the test group spend studying the M-Learning materials in order to attempt an estimate of the amount of time expended relative to any assessment measure (e.g., minutes studied per test score percent, or similar). Such a measure might elucidate the nature of the relationship between the effectiveness of M-Learning materials with respect to the more labor intensive live lecture alternative. Future work should also consider gender and GPA between the study groups, to control for possible confounding effects of those parameters on test results.

When deployed via handheld mobile devices or over 3G networks, specialized content such as presented in this course can offer novel, flexible education or professional certification maintenance opportunities for those distant to a college campus (e.g., deployed military, overseas, or rural locations), or having significant commuter or job-related travel time.

Acknowledgment

The author wishes to thank Ms. Katie Hackett for her work in preparing the references for the introductory section of this work.

REFERENCES

[1] S. Wexler, B. Schlenker, J. Brown, and D. Metcalf, "Guild Research 360 Report on Mobile Learning." The eLearning Guild Research, 7 August 2007. Available from http://www.elearningguild.com/research/archives/index.cfm?action=view&frompage=1&StartRow=1&MaxRows=40&selection=doc.30, last accessed October 12, 2008.

[2] C. O'Malley, G. Vavoula, J. Glew, J. Taylor, M. Sharples, and P. Lefrere, "Guidelines for learning/teaching/tutoring in a mobile environment." Mobilearn project deliverable, 2003. Available from http://www.mobilearn.org/download/results/guidelines.pdf, last accessed October 12, 2008.

[3] M. Sharples, "A Short History of Mobile Learning." History of Mobile Learning, 2007. The University of Nottingham. Available from http://seriousgames.ning.com/group/mobilelearning/forum/topic/show?id=630751%3ATopic%3A22834, last accessed October 12, 2008.

[4] P. Rainger, "Accessibility and M-learning." Notes from the Joint Accessibility and EC (Educational Content) SIG Meeting, 23 May 2005. CETIS. Available from http://zope.cetis.ac.uk/members/accessibility/meetings/2005/sig12/overview, last accessed October 12, 2008.

[5] D. Keegan, "The Incorporation of Mobile Learning into Mainstream Education and Training." 2005. Available from http://www.mlearn.org.za/CD/papers/keegan1.pdf, last accessed October 12, 2008.

[6] J. Kossen, "When E-learning Becomes M-Learning." Computing Unplugged. Available from http://www.palmpowerenterprise.com/issues/issue200106/elearning001.html, last accessed October 12, 2008.

[7] J. Traxler, "Landscape Study on the Use of Mobile and Wireless Technologies for Learning and Teaching in the Post-16 Sector." Notes from the Joint Accessibility and EC (Educational Content) SIG Meeting. 20 May 2008. Available from http://zope.cetis.ac.uk/members/accessibility/meetings/2005/sig12/overview, last accessed October 12, 2008.

[8] Joint Information Systems Committee, "Introduction to M-learning." JISC TechDis. Available from http://www.techdis.ac.uk/index.php?p=9_5_32_1, last accessed October 12, 2008.

[9] L. Motiwalla, "Mobile Learning: A Framework and Evaluation." *Computers and Education*, vol. 49, pp. 581-596, 2007.

[10] P. McCloskey**,** *"GT Prof: Students Learn Better Via iPod Versus Lecture,"* Campus Technology, 13 Mar 2007, Available from http://www.campustechnology.com/article.aspx?aid=45383, last accessed October 12, 2008.

[11] W. McNairy, "Spring 2007 iPod Courses." 16 Aug 2007. Duke University Center for Instructional Technology. 11 May 2008. Available from http://cit.duke.edu/help/ddi/archive_ipod_spring2007.html#193, last accessed October 12, 2008.

[12] M. Prensky, "What can you learn from a Cell Phone? Almost Anything!" *Innovate,* vol. 1, June/July, 2005. Available from http://www.innovateonline.info/index.php?view=article&id=83, last accessed October 12, 2008.

[13] M. Kato, and V. Ricci, "Mobile Learning in Japan: Why the Future has Already Arrived in Asia." *The eLearning Guild's Learning Solutions,* pp. 1-11, September, 2006. Available from http://www.elearningguild.com/pdf/2/091106mgt-h.pdf, last accessed October 12, 2008.

[14] J. Marshall, "Smartphones are the PCs of the Developing World." NewScientistTech, 1 Aug 2007. Available from http://technology.newscientist.com/channel/tech/mg19526156.300-smartphones-are-the-pcs-of-the-developing-world.html, last accessed October 12, 2008.

A Simplified Array Configuration Based GSC-SSF Method

Zhang Cong, He Xinghua, Chang Qing, Lu Huanzhang
School of Electronic Science and Engineering National Univ. of Defense Technology Changsha, Hunan 410073, P.R. China
huaxinghe@nudt.edu.cn

***Abstract*-A new simplified array configuration based method for estimating the two-dimensional (2-D) direction-of-arrival (DOA) of cyclostationary signals is presented. By eigen-decomposing the pseudo-data Matrices constructed from the cyclic correlation functions of sensor outputs, the proposed method can be applied to not only the wideband but also the narrowband signals. Moreover, the 2-D DOAs can be estimated without 2-D searching or parameter pairing procedures. Computer simulations demonstrate the capabilities of our method.**

I. INTRODUCTION

In recent years, the cyclic approach has been widely applied to source location problems [1-4]. By exploiting the second-order cyclostationarity properties exhibited by most communication signals, cyclic methods can null out or greatly reduce the effect of other co-band interferences and background noises. Hence, better performances can be achieved when comparing with conventional methods. Moreover, by manipulating the cyclic correlation functions of sensor outputs, the temporal delay of signal can be transformed into a phase shift of the corresponding cyclic correlation function regardless of the bandwidth of the signal [5]. Using this property, several cyclic methods for wideband signal direction-of-arrival (DOA) estimation were proposed, such as [5] and [6].

However, these methods can only resolve directions of sources located in a plane. When comes to the problem of two-dimensional (2-D) DOA estimation for wideband signals, little work can be found. Ref. [7] presents a 2-D generalized spectral correlation signal subspace fitting (GSC-SSF) method utilizing the dual parallel linear array. By constructing and eigen-decomposing a so-called DOA matrix, this method avoids 2-D searching and parameter pairing procedures. Whereas, the two identical subarrays configuration used by GSC-SSF method is costly in hardware and computation [8]. Accordingly, Ref. [8] presents a simplified array configuration, based on which the same performance as that in [7] can be achieved, but with much fewer sensors. The method in [8] uses cyclic auto-correlation function to achieve the estimation of 2-D DOA. However, we notice that cyclic cross correlations of different sensor outputs performs better than cyclic auto-correlations of the same sensor outputs in noise suppressing due to the limited observations obtained in some real environments. Thus, a cyclic cross-correlation function of neighboring sensor outputs based method for wideband cyclostationary signals is proposed in our previous work [9].

In this paper, a generalized simplified array configuration based GSC-SSF method is proposed. Comparing with the dual parallel array used by GSC-SSF method, the proposed simplified array requires fewer sensors and accordingly reduces sensor matching problems in parallelism and consistency that would be encountered in practical applications. By properly exploiting the cyclic correlation functions of sensor outputs, the proposed method constructs two rotational invariable "pseudo-data" matrices, based on which the DOA matrix is constructed. Then, the DOAs can be estimated by eigen-decomposing the DOA matrix. Moreover, this paper indicates that both of the method in [8] and [9] are just special cases of the proposed generalized method.

The paper is organized as follows. Section II presents a basic review of the concepts of cyclostationarity. In section III, the simplified array configuration is introduced first, and then an appropriate mathematical model on the basis of cyclic correlation function is constructed. In section Ⅳ and Ⅴ, we propose a new method for estimating 2-D DOA with specific steps and discuss the cases when sensor number degrades. Finally, several computer simulations are conducted aiming at evaluating performances of the proposed method, and then the paper is concluded.

II. CYCLOSTATIONARITY

A random process $s(t)$ is called wide-sense cyclostationary if its mean $m_s(t)=E\{s(t)\}$ and auto-correlation $R_{ss}(t,u)=E\{s(t)s^*(u)\}$ are periodic with some period, say $1/\alpha$:

$$m_s(t+k/\alpha)=m_s(t) \tag{1}$$

$$R_{ss}(t+k/\alpha,\, u+k/\alpha)=R_{ss}(t,u) \tag{2}$$

where k is an integer and the asterisk denotes complex conjugate. α is cycle frequency. Generally speaking, different signals have different cycle frequency set $\{\alpha\}$. By selecting an appropriate α, we can extract the signals of interest (SOIs) out of other signals that do not exhibit cyclostationarity at α. This ability is sometimes referred to as signal selectivity.

The cyclic correlation function and conjugate cyclic correlation function of $s(t)$ are defined as [5]

$$R_{ss}^{\alpha}(\tau)=\left\langle s(t+\tau/2)s^*(t-\tau/2)\exp(-j2\pi\alpha t)\right\rangle_t \tag{3}$$

$$R_{ss^*}^{\alpha}(\tau)=\left\langle s(t+\tau/2)s(t-\tau/2)\exp(-j2\pi\alpha t)\right\rangle_t \tag{4}$$

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_18, © Springer Science+Business Media B.V. 2010

where $\langle g \rangle_t = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} g\, dt$.

Lemma [5]**:** if $s(t)$ is a cyclostationary process with cyclic auto-correlation function $R_{ss}^{\alpha}(\tau)$, and $s'(t) = s(t+T)$, then

$$R_{s's'}^{\alpha}(\tau) = R_{ss}^{\alpha}(\tau)\exp(j2\pi\alpha T) \tag{5}$$

We should notice that $R_{ss}^{\alpha}(\tau)$ is not a function of t and the time delay is transformed into a phase shift of its cyclic auto-correlation function.

III. Cyclic Correlation based Data Model

Consider a sensor array system consisting of one uniformly spaced linear subarray X_a and 4 guiding sensors, as shown in Fig. 1. X_a lies on the X-AXIS and starts at the origin. It has M isotropic sensors and D is the separation of neighboring sensors. The locations of the guiding sensors are (D_g,Δ) , $(D_g + D,\Delta)$, $(D_g + lD,\Delta)$ and $(D_g + (l+1)D,\Delta)$, respectively, where l is a given order for the proposed generalized array geometry. l is a integer and $0 \le l \le M-1$. Assume that K sources impinge on the array from directions $\theta_1,\theta_2,\cdots,\theta_K$, where $\theta_k = \{\alpha_k,\beta_k,\gamma_k\}$ and $\alpha_k,\beta_k,\gamma_k$ is the DOA of the kth source relative to X-axis, Y-axis and Z-axis, respectively.

As γ_k can be given by

$$\cos^2\alpha_k + \cos^2\beta_k + \cos^2\gamma_k = 1 \tag{6}$$

only (α_k,β_k) needs to be estimated in what follows.

The output of the ith sensor in subarray X_a can be expressed as

$$x_i(t) = \sum_{k=1}^{K} s_k(t + (i-1)\tau_{Dk}) + n_i(t), \quad i = 1,2,\ldots,M \tag{7}$$

where $\tau_{Dk} = D\cos\alpha_k / c$ with c as the wave speed; $s_k(t)$ is the kth signal impinging on the reference sensor (sensor at the origin); $n_i(t)$ is the noise at the ith sensor in X_a .

Similarly, the outputs of the guiding sensors are

$$g_j(t) = \sum_{k=1}^{K} s_k(t + (\frac{D_g}{D} + j - 1)\tau_{Dk} + \tau_{\Delta k}) + n_{gj}(t), \quad j = 1,2 \tag{8}$$

$$g_j(t) = \sum_{k=1}^{K} s_k(t + (\frac{D_g}{D} + l + j - 3)\tau_{Dk} + \tau_{\Delta k}) + n_{gj}(t), \quad j = 3,4 \tag{9}$$

where $\tau_{\Delta k} = \Delta\cos\beta_k / c$; $n_{gj}(t)$ is the noise at the jth guiding sensor.

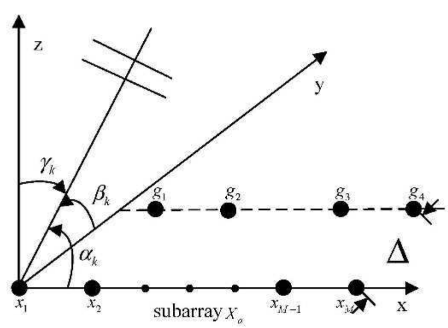


Fig. 1. Proposed simplified array.

Without loss of generality, assume that there are only d ($d \le K$) signals exhibit second-order cyclostationarity (i.e., they are SOIs) with a common cycle frequency α . Other signals that either have different cycle frequencies or do not exhibit cyclostationarity are considered as interferences. In addition, the following assumptions are made:

a) The SOI's are mutually cyclically uncorrelated [5], that is, any pair of SOI's does not exhibit joint second-order cyclostationarity with the considered cycle frequency α .
b) The interfering signals and the noises are cyclically uncorrelated with the SOIs.
c) The interfering signals and the noises are mutually cyclically uncorrelated at the considered cycle frequency α .

Under assumptions a)-c), the cyclic cross-correlation function $R_{x_p x_q}^{\alpha}(\tau)$, $p,q = 1,2,\ldots,M$ in X_a is given by

$$R_{x_p x_q}^{\alpha}(\tau) = \left\langle x_p(t+\tau/2)x_q^*(t-\tau/2)\exp(-j2\pi\alpha t)\right\rangle_t$$
$$= \sum_{k=1}^{d} R_{s_k s_k}^{\alpha}(\tau + (p-q)\tau_{Dk})\exp(j\pi\alpha(p+q-2)\tau_{Dk}) \tag{10}$$

The proof of the above equation can be found in Ref. [7].

Accordingly, it is easy to obtain $R_{x_{q+l}x_q}^{\alpha}(\tau), q = 1,\ldots,M-l$ as

$$R_{x_{q+l}x_q}^{\alpha}(\tau) = \left\langle x_{q+l}(t+\tau/2)x_q^*(t-\tau/2)\exp(-j2\pi\alpha t)\right\rangle_t$$
$$= \sum_{k=1}^{d} R_{s_k s_k}^{\alpha}(\tau + l\tau_{Dk})\exp(j\pi\alpha(2q+l-2)\tau_{Dk}) \tag{11}$$

Similarly, we can obtain $R_{g_{i+2}g_i}^{\alpha}(\tau)$, $i = 1,2$ as

$$R_{g_{i+2}g_i}^{\alpha}(\tau) = \left\langle g_{i+2}(t+\tau/2)g_i^*(t-\tau/2)\exp(-j2\pi\alpha t)\right\rangle_t$$
$$= \sum_{k=1}^{d} R_{s_k s_k}^{\alpha}(\tau + l\tau_{Dk})\exp(j\pi\alpha((2\frac{D_g}{D} + l + 2(i-1))\tau_{Dk} + 2\tau_{\Delta k})) \tag{12}$$

Collect the above functions in two vectors as follows

$$\mathbf{R}_1^{\alpha}(\tau) = \left[R_{g_3 g_1}^{\alpha}(\tau), R_{x_{1+l}x_1}^{\alpha}(\tau),\ldots,R_{x_{M-1}x_{M-1-l}}^{\alpha}(\tau)\right]^T \tag{13}$$

$$\mathbf{R}_2^{\alpha}(\tau) = \left[R_{g_4 g_2}^{\alpha}(\tau), R_{x_{2+l}x_2}^{\alpha}(\tau),\ldots,R_{x_M x_{M-l}}^{\alpha}(\tau)\right]^T \tag{14}$$

where $[\mathrm{g}]^T$ denotes transpose. If we define steering matrix $\mathbf{A}^{\alpha}$ as

$$\mathbf{A}^{\alpha} = \left[\mathbf{a}_1^{\alpha},\mathbf{a}_2^{\alpha},\ldots,\mathbf{a}_d^{\alpha}\right] \tag{15}$$

where $\mathbf{a}_k^{\alpha}$ for $k = 1,\ldots,d$ is defined as

$$\mathbf{a}_k^{\alpha} = [\exp(j\pi\alpha(2\frac{D_g}{D}\tau_{Dk} + l\tau_{Dk} + 2\tau_{\Delta k})), \exp(j\pi\alpha l\tau_{Dk}),$$
$$\exp(j\pi\alpha(l+2)\tau_{Dk}),\cdots,\exp(j\pi\alpha(2M-l-4)\tau_{Dk})]^T \tag{16}$$

and define diagonal matrix $\mathbf{\Phi}^{\alpha}$ as

$$\mathbf{\Phi}^{\alpha} = diag\left[\exp(j2\pi\alpha\tau_{D1}),\exp(j2\pi\alpha\tau_{D2}),\ldots,\exp(j2\pi\alpha\tau_{Dd})\right] \tag{17}$$

equation (13) and (14) can be rewritten as

$$\mathbf{R}_1^{\alpha}(\tau) = \mathbf{A}^{\alpha}\mathbf{R}_S^{\alpha}(\tau) \tag{18}$$

$$\mathbf{R}_2^{\alpha}(\tau) = \mathbf{A}^{\alpha}\mathbf{\Phi}^{\alpha}\mathbf{R}_S^{\alpha}(\tau) \tag{19}$$

where $\mathbf{R}_S^{\alpha}(\tau) = \left[R_{s_1 s_1}^{\alpha}(\tau + l\tau_{D1}), R_{s_2 s_2}^{\alpha}(\tau + l\tau_{D2}),\ldots,R_{s_d s_d}^{\alpha}(\tau + l\tau_{Dd})\right]^T$ is the cyclic auto-correlation vector of SOIs.

From (18) and (19), we notice that these data models can be obtained regardless of the signal bandwidth. That is to say, these data models can be applied to either narrowband or wideband sources.

Since all the DOAs of SOIs are involved in $\mathbf{A}^{\alpha}$ and $\mathbf{\Phi}^{\alpha}$, estimation of $\mathbf{A}^{\alpha}$ and $\mathbf{\Phi}^{\alpha}$ out of $\mathbf{R}_1^{\alpha}(\tau)$ and $\mathbf{R}_2^{\alpha}(\tau)$ becomes the problem waiting to be resolved. Considering that $\mathbf{R}_1^{\alpha}(\tau)$ and $\mathbf{R}_2^{\alpha}(\tau)$ are not matrices but vectors, we present a method exploiting the cyclic correlation function of multiple time delays in the following discussions in order to achieve more accurate results.

IV. THE PROPOSED 2-D DOA ESTIMATION METHOD

Sample $\mathbf{R}_1^{\alpha}(\tau)$ and $\mathbf{R}_2^{\alpha}(\tau)$ uniformly at different lags τ and take them as "snapshots", then obtain "pseudo-data" matrices as follows

$$\mathbf{X}^{\alpha}=\left[\mathbf{R}_1^{\alpha}(0),\mathbf{R}_1^{\alpha}(T_s),\ldots,\mathbf{R}_1^{\alpha}((N-1)T_s)\right] \tag{20}$$

$$\mathbf{Y}^{\alpha}=\left[\mathbf{R}_2^{\alpha}(0),\mathbf{R}_2^{\alpha}(T_s),\ldots,\mathbf{R}_2^{\alpha}((N-1)T_s)\right] \tag{21}$$

where T_s is the sampling period and N is the pseudo snapshots.

From (18) and (19), the above equations can be rewritten as

$$\mathbf{X}^{\alpha}=\mathbf{A}^{\alpha}\mathbf{S}^{\alpha} \tag{22}$$

$$\mathbf{Y}^{\alpha}=\mathbf{A}^{\alpha}\mathbf{\Phi}^{\alpha}\mathbf{S}^{\alpha} \tag{23}$$

where $\mathbf{S}^{\alpha}=\left[\mathbf{R}_S^{\alpha}(0),\mathbf{R}_S^{\alpha}(T_s),\ldots,\mathbf{R}_S^{\alpha}(NT_s)\right]$ is the pseudo signal matrix. From the expressions of $\mathbf{X}^{\alpha}$ and $\mathbf{Y}^{\alpha}$, it is easy to find that $\mathbf{X}^{\alpha}$ and $\mathbf{Y}^{\alpha}$ can be seen as the outputs of two parallel subarrays with rotational invariance property. In such a case, we can construct the DOA matrix and consequently estimate the DOAs by eigen-decomposing it.

Define DOA matrix as

$$\mathbf{R}^{sc}=\mathbf{Y}^{\alpha}\left[\mathbf{X}^{\alpha}\right]^{-} \tag{24}$$

where $[\cdot]^{-}$ denotes pseudo-inverse. The following theorem provides the foundation for the results presented herein.

Theorem : If both $\mathbf{A}^{\alpha}$ and $\mathbf{S}^{\alpha}$ are full-rank and all the diagonal elements of $\mathbf{\Phi}^{\alpha}$ are different, the d nonzero eigenvalues of $\mathbf{R}^{sc}$ equal to the d diagonal elements of $\mathbf{\Phi}^{\alpha}$ and the corresponding eigenvectors equal to the d column vectors of matrix $\mathbf{A}^{\alpha}$, i.e.,

$$\mathbf{R}^{sc}\mathbf{A}^{\alpha}=\mathbf{A}^{\alpha}\mathbf{\Phi}^{\alpha} \tag{25}$$

The proof is similar to that in [10]. Thus, we can obtain $\mathbf{A}^{\alpha}$ and $\mathbf{\Phi}^{\alpha}$ by eigen-decomposition $\mathbf{R}^{sc}$, from which many ways can be used to estimate (α_k,β_k). For example, α_k can be estimated through kth nonzero eigenvalue of $\mathbf{R}^{sc}$, and β_k can be estimated through the corresponding eigenvector of $\mathbf{R}^{sc}$. Notice that α_k and β_k are calculated through the corresponding eigenvalue and eigenvector of $\mathbf{R}^{sc}$. Therefore, α_k and β_k are paired automatically.

Summarize the proposed algorithm as follows:

Step1: For a fixed l, the sensor array configuration is achieved;

Step2: For signals received at the sensor array, and for specified cycle frequency α (estimated or known) estimate the cyclic cross-correlation $R^{\alpha}_{x_{p+l}x_p}(\tau)$ and $R^{\alpha}_{g_{j+2}g_j}(\tau)$, where $\tau=0,T_s,\ldots,NT_s$ and $p=1,2,\ldots,M-l$, $j=1,2$;

Step3: Form pseudo-data matrices $\mathbf{X}^{\alpha}$ and $\mathbf{Y}^{\alpha}$;

Step4: Calculate the DOA matrix $\mathbf{R}^{sc}$;

Step5: Apply eigen-decomposition to $\mathbf{R}^{sc}$, and then estimate the 2-D DOAs of all the SOIs, i.e., (α_k,β_k), $k=1,2,\ldots,d$, by eigenvalues and the corresponding eigenvectors of $\mathbf{R}^{sc}$.

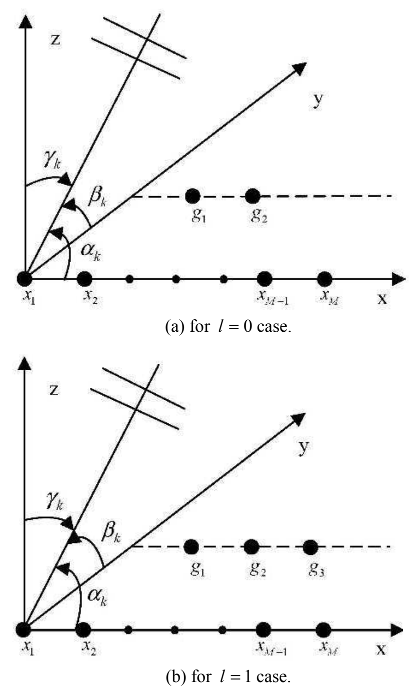
V. DISCUSSIONS

A. Number of Guiding Sensors

As can be seen from the above discussions and Fig.1, for a fixed l, the sensor array configuration and the corresponding signal processing models are confirmed. However, the guiding sensor number degrades when $l=0,1$. Thus, the corresponding two cases are discussed below.

Case 1: When $l=0$, sensor g_{i+2} and g_i $(i=1,2)$ locate in the same place. Thus, the guiding sensor number degrades to 2. Correspondingly, $R^{\alpha}_{x_{q+l}x_q}(\tau)$ and $R^{\alpha}_{g_{i+2}g_i}(\tau)$ in (11) and (12) turn into the cyclic auto-correlation function $R^{\alpha}_{x_qx_q}(\tau)$ and $R^{\alpha}_{g_ig_i}(\tau)$. Observing the array configuration in Fig.2 (a) and the signal processing models in this case, it is easy to find that the proposed method with $l=0$ is just the same as the DOA method presented in [8]. That is to say, the DOA method in Ref. [8] is just a special example of the proposed method.

Case 2: When $l=1$, sensor g_2 and g_3 locate in the same place. Thus, the guiding sensor number degrades to 3. The corresponding array configuration can be seen in Fig.2 (b). Here, the cyclic correlation functions in (11) and (12) are



(a) for $l=0$ case.

(b) for $l=1$ case.

Fig. 2 Array configuration.

cyclic cross-correlation functions of neighboring sensor outputs. Similar to case 1, it is also easy to find that the proposed method with $l=1$ is just the same as the DOA method presented in [9].

B. The Proposed Method Using Conjugate Cyclostationarity of SOIs

Many cyclostationary signals, e.g. BPSK, AM also exhibit conjugate cyclic property [11] (i.e., the conjugate cyclic correlation is nonzero for some cycle frequency α). Hence, we can also apply this property to process the signals.

Similar to (11) and (12), the conjugate cyclic correlation function $R^{\alpha}_{x_{q+l}x_q^*}(\tau)$, $p=1,\ldots,M-l$ and $R^{\alpha}_{g_{i+2}g_i^*}(\tau)$, $i=1,2$ can be expressed as

$$R^{\alpha}_{x_{q+l}x_q^*}(\tau)=\sum_{k=1}^{d}R^{\alpha}_{s_k s_k^*}(\tau+l\tau_{Dk})\exp(j\pi\alpha(2q+l-2)\tau_{Dk}) \tag{26}$$

$$R^{\alpha}_{g_{i+2}g_i^*}(\tau)=\sum_{k=1}^{d}R^{\alpha}_{s_k s_k^*}(\tau+l\tau_{Dk})\exp(j\pi\alpha((2\frac{D_g}{D}+l+2(i-1))\tau_{Dk}+2\tau_{\Delta k})) \tag{27}$$

Replace $R^{\alpha}_{x_{q+l}x_q}(\tau)$ and $R^{\alpha}_{g_i g_i}(\tau)$ in (13) and (14) with $R^{\alpha}_{x_{q+l}x_q^*}(\tau)$ and $R^{\alpha}_{g_{i+2}g_i^*}(\tau)$ respectively. Then, same procedures are adopted as the proposed cyclic DOA estimation and the DOAs can be estimated thereafter. In the following examples of computer simulation, the conjugate cyclic properties of the signals are exploited.

VI. Computer Simulation Examples

Computer simulations for testing the DOA estimation ability of the proposed method are conducted. And the methods utilized here are based on the conjugate cyclostationarity of the signals received.

Consider the proposed simplified array consisting of a M -element subarray and 4 guiding sensors as shown in Fig.1. The cyclic auto-correlation matrix of the complex additive white Gaussian noise (AWGN) equals to $\sigma^2\boldsymbol{I}$ for cycle frequency 0 and zero for others. Unless otherwise specified, the statistics of the DOA estimates were calculated based on 400 independent trials.

In the first simulation, the accuracy of the proposed method is tested. Assume that one wideband SOI arrives from $(65^{\circ},80^{\circ})$ and no interferences are present. The SOI is a BPSK signal with baud rate $f_b=1$ Mb/s and carrier frequency $f_0=3$ MHz. Since BPSK signal exhibit second-order conjugate cyclostationarity for $2f_0+kf_b$, the cycle frequency is set as $2f_0$. The sample number is 4096 and the pseudo snapshots are 400. Fig.3 and Fig.4 show how the signal-to-noise ratio (SNR) affects the 2-D DOA estimation results of the proposed method with $l=0$ and $l=1$ respectively. The performance of the method is quantified by the root mean square error (RMSE). As can be seen from the simulation results, the proposed method with $l=1$ provides more accurate estimates than the method with $l=0$. Moreover, as the subarray number M rises from 7 to 10, both of the two methods improve their α_k estimation performance, but almost help nothing about the β_k estimation ability.

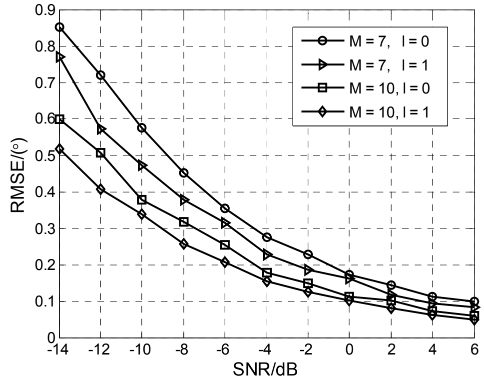


Fig. 3 RMSE of α_k estimates versus SNR.

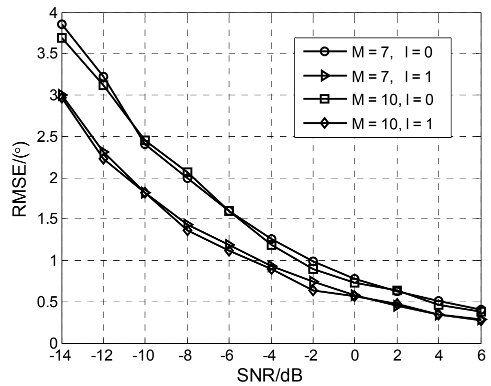


Fig. 4 RMSE of β_k estimates versus SNR.

Under the same condition as the first simulation, the second simulation is conducted to test the effect of sample number on DOA estimation. Subarray number M is set to 10, and SNR= −10 dB. Pseudo snapshots are 400. Fig.5 and Fig.6 show that the proposed method provides more accurate estimates in both α_k and β_k as the sample number grows.

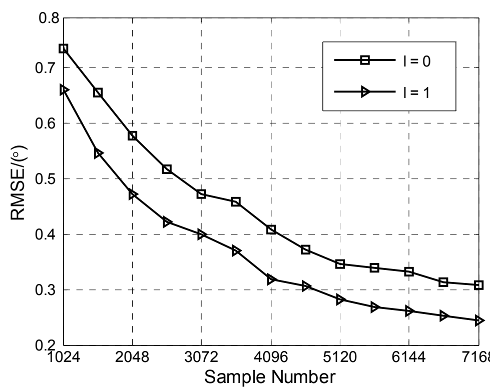


Fig. 5 RMSE of α_k estimates versus Sample Number.

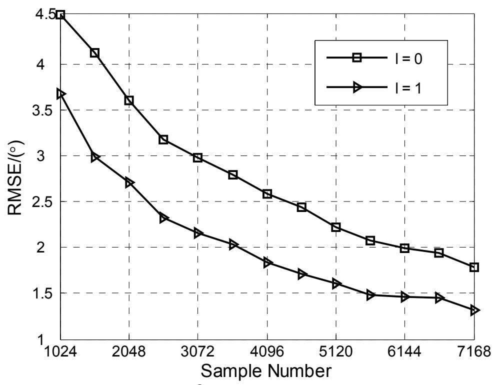


Fig. 6 RMSE of β_k estimates versus Sample Number.

In the third simulation, signal selectivity ability of the proposed method is verified. Assume that two signals impinge on the sensor array. One is a SOI arriving from $(65^\circ, 80^\circ)$ and the other is a interference arriving from $(80^\circ, 120^\circ)$. Both of them have the same carrier frequency f_0. Only the SOI exhibits conjugate cyclostationarity at $\alpha = 2f_0$. M is set to 10 and SNR= - 5 dB. The sample number is 4096 and the pseudo snapshots are 400. Fig. 7 shows the joint DOA estimation results of the proposed method with $l=1$ in 200 independent trials. As can be seen from Fig. 6, the proposed method only estimates the 2-D DOA of the SOI and suppresses the interference. This simulation indicates that the proposed method provides good signal selectivity.

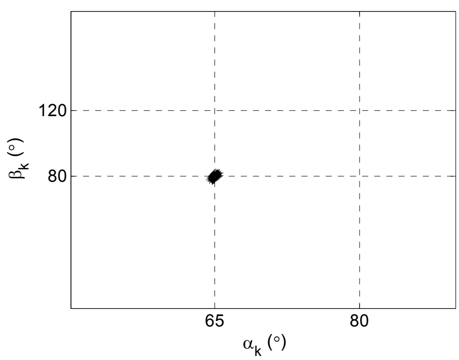


Fig. 7 2-D DOA estimation results.

VII. Conclusions

In this paper, we have presented a generalized simplified array configuration based 2-D DOA estimation method. This method combines temporal and spatial properties effectively, which makes the applicability of the method to either narrowband or wideband signals. Theoretical analysis and simulation results proved that the proposed method has better signal selectivity and better performance in noise suppressing. Furthermore, the proposed method does not require any 2-D search and the 2-D parameters can be automatically paired, which means the problem of large computation can be circumvented.

Acknowledgment

The authors would thank the anonymous reviewers gratefully for their stimulating comments, which helped improve the quality of the paper.

References

[1] H. EI Markhi, M. M. O. Haibala, F. Mrabti, P. Charge and M. Zouak, “An improved cyclic beamforming method for signal DOA estimation,” Signal, Image and Video Processing, vol. 1: 267–272, August 2007.

[2] P. Charge, Y. Wang and J. Saillard, “An extended cyclic MUSIC algorithm,” IEEE Trans. Signal Process. , 51(7): 1695-1701, July 2003.

[3] G. Gelli and L. Izzo, “Cyclostationarity-based coherent methods for wideband-signal source location,” IEEE Trans. On Signal Processing. 51(10): 2471-2482, Oct. 2003.

[4] W. A. Gardner, “Simplification of MUSIC and ESPRIT by exploitation of cyclostationarity,” Proc. of IEEE, 76(7): 845-847, July 1988.

[5] G. H. Xu and T. Kailath, “Direction-of-arrival estimation via exploitation of cyclostationarity-A combination of temporal and spatial processing,” IEEE Trans. Signal Process., 1992, 40(7): 1775-1786.

[6] H. Yan and H. H. Fan, “Wideband cyclic MUSIC algorithms,” Signal Processing, vol. 85: 643-649, March 2005.

[7] Liang Jin, Qinye Yin and Bofeng Jiang, “Direction Finding of Wideband Signals via Spatial-Temporal Processing in Wireless Communications,” IEEE international Symposium on Circuits and Systems, ISCAS’00, Geneva, Switzerland, vol. 2: 81-84, 2000.

[8] Liang Jin, Wenjie Wang and Qinye Yin, “Two-dimensional Direction Finding for Wideband Signals Using Simplified Array Configuration,” Proc. of IEEE ISCAS, vol. 3: 78-81, 1999.

[9] Cong Zhang, Deli Chen, Huanzhang Lu, Huamin Tao. Cyclic Cross-correlation based Two-dimensional Direction of Arrival Estimation Algorithm. IEEE International Conference on Information and Automation (ICIA’08), 2008.6. pp.1148-1152.

[10] L. Jin, *Time-space eigenstructure based array signal processing with smart antenna system*, ph. D Thesis of Xi’an Jiaotong University, Xi’an, 1999.

[11] Jia, W.M., Yao, M.L. and Song, J.S.. “Spectral correlation MUSIC algorithm for cyclostationary signals by exploitation of new array configuration”. IEEE International Symposium on ISCAS, 4(5): 3159-3162, 2005.

"UniTeSys" - High School Student Knowledge Assessment Tool

Borislav Slavov
bslavov@uctm.edu

Abstract—**The testing system "UniTeSys" is designed for assessment of student's knowledge in whatever discipline of high school curriculum. The determination of the field's discipline is given by generating of suitable database, containing questions, answers and pictures for the curriculum material. The dialog between the system and the tested student is clear and precludes any misunderstandings. The system use allows the tested studying material and the visual environment both to be introduced in Latin alphabet (English) or in Cyrillic alphabet (Bulgarian). The system gives the opportunity of group test in small computer labs up to 10 machines in different disciplines. Using the system by the customers doesn't require any special hardware and software resources. Fairly enough is the use of a computer system with operating system Windows 95 and above.**

Index Terms— **test system, assessment of knowledge, teachers training tools**

I. INTRODUCTION

Test systems play an important role in the educational process. Therefore, many researchers are involved in the development, improvement and evaluation of the test computer systems [1], [2], [3]. Bulgarian primary and high education is under active absorption of new technologies in IT, there are built modern computer labs with Internet. Bulgarian teachers have to apply in practice modern training methods and new ways of assessing students' knowledge. Lack of adequate and useful tools for the work of the teacher set to be a tool for evaluating test knowledge of students.
Requirements for this tool are:

• Compatible with populate operating systems in school labs;
• To operate in LAN or WiFi networks;
• Independent of computer hardware;
• Easy and friendly interface;
• To enjoy the wide range of users;
• Small volume;
• Flexible methods of evaluation;

To achieve the objective were the following tasks formulated to this tool:

• To create an appropriate architecture;
• To design clear, easy and quick User Interface;
• To design and implement appropriate and flexible modules for objective evaluation of the student's knowledge;
• To analyze the assessment of student knowledge with appropriate statistical module;

II. SPECIFICATIONS

The testing system "UniTeSys" obtains the following parameters:

Table 1
Specifications of the testing system "UniTeSys"

Parameter	Description
Total number of questions in a test	300
Total number of the TRUE answers	defined by the teacher
Total number of the FALSE answers	defined by the teacher
Total number of attached pictures(*.bmp)	defined by the teacher
Total signs (ASCII format) per row	120
Evaluate (mark)	scale: from (2.00) to (6.00)
	points
Assessment methods	depends of the number of given answers
	depends of the weights of the given answers
Final evaluation of the test	average
	scale

III. DESCRIPTION

The testing system "UniTeSys" is designed and developed as "all-in-one" single application (executable file). It includes:

- Application authorization module;
- Database files Crypt/Decrypt module;
- Student test files Crypt/Decrypt module;
- Built in own RDBMS(Relation Database Management System)
- Built in small statistical module;
- Simple and friendly user interface;
- Hardware undependable(no matter slow, fast machine or special requirements);
- Small single executable file(649kb);
- Usage of ordinary ASCII text file format for input of questions and answers and student test results;
- Usage of commonly popular graphical file format for pictures(BMP);
- Flexible assessment methods module;

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_19, © Springer Science+Business Media B.V. 2010

- 2 access profiles: student and teacher;
- Built in short user guide and assessment methods description module;

The next figure illustrates the architecture of the testing system:

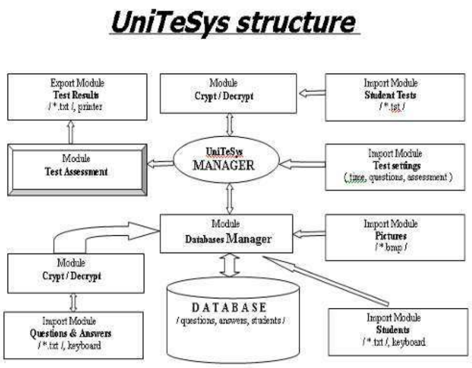


figure 1

IV. ASSESSMENT

For assessing the answers to separate questions included in the test the following approach was chosen:

P A R T - I
Formation of the mark of the separate question

The testing system “UniTeSys” offeres two methods of question mark formation based on the given answers.

M E T H O D - 1
Without answers weight coefficients

Resume:
This method assumes that answers in the question have equal weights, so their count is important only. On base of given TRUE answers to the questions from,a mark is calculated by interpolation in range of 2.00 and 6.00 on chosen mark's scale / level 0 /. In case that FALSE answers are given, reduction of calculated mark is done / level 1 /.

Description:
The mark of the question is calculated at two stages:

STAGE 1
Formation of the base mark of the question
("the mark of knowledge")
| level 0 |

Let's suggest that (x) is the number of given TRUE answers by the student to the question. If the inequality $x \geq x1$ is done (the level of knowledge is achieved), then we assumed that student has successfully answered the question and the testing system calculates a base mark by interpolation using formula of equation of line with the following parameters:

x1 - compulsory level of knowledge(it is forming as % from x2);
x2 - total number of all TRUE answers included into the question;
y2 - maximal mark in the test by chosen scale;
y1 - minimal mark in the test by chosen scale;
a, b - coefficients in the formula of equation of line;
x - number of given TRUE answers by the student;
O1(x) - calculated base mark of the question;

STAGE 2
Reduction of the calculated base mark of the question
("the real mark")
| level 1 |

If there are no FALSE answers (x1=0) then the final mark is equal of calculated base mark O1(x) by STAGE I. In the opposite case the calculated base mark O1(x) is reduced to mark O2(x) using formula of equation of line as reduction equation with the following parameters:

x1 - minimal number of given FALSE answers in the question;
x2 - total number of all FALSE answers included into the question;
y2 - minimal mark of the test;
y1 - calculated base mark O1(x) by STAGE 1;
a,b - coefficients into the formula of reduction equation of line;
x - number of given FALSE answers from the student;
O2(x) - calculated (reduced) final mark of the question;

M E T H O D - 2
With answers weight coefficients

Resume:
In this method the tutor in advance sets correct weights of separate answer in the question such as:
- the TRUE answer obtains weight coefficient > 0.00
- the FALSE answer obtains weight coefficient < 0.00
Calculates a sum of weights of given answers in the question. Then by interpolation on chosen scale calculates a mark in frame of marks between [2.00 - 6.00]
/ level 2 /. The calculated mark is final and there is no need of its reduction.

Description:
Formation of the mark of the question
("the mark of knowledge")
| level 2 |

Let's assume that (x) is the sum of weights of answers to the question. If the inequality $x \geq x1$ is done (the level of knowledge is achieved), then it is assumed that student has successfully answered the question and the testing system calculates a mark by interpolation using formula of equation of line with the next parameters:

x1 - compulsory level of knowledge(it is forming as % from x2);

x2 - sum of weights of all TRUE answers included in the question;

y2 - maximal mark in the test by chosen scale;

y1 - minimal mark in the test by chosen scale;

a, b - coefficients in the formula of equation of line;

x - sum of weights of given answers by the student;

O1(x) - calculated mark of the question;

P A R T - II
Formation of the final mark of the test

The testing system "UniTeSys" offers two methods of the final mark formation:

M E T H O D - 1
Average mark

Resume:
In this method the final mark is calculated as sum of all marks of the separated questions in the student test and is divided on the number of the questions.

M E T H O D - 2
Mark by scale

Resume:
Let's suggest that (x) is the sum of all marks ≥3.00. On the base of this sum is calculated a base mark of the whole test by chosen scale. If there are marks of the questions less than 3.00 then reduction is made of the test base mark.

Description:
The final mark of the test is calculated in two stages:

STAGE 1
Formation of the base mark of the whole test
("the mark of knowledge")

Let's suppose that (x) is the sum of all marks in the range of marks between 3.00 and 6.00. If the inequality x=>x1 is done (the level of knowledge is achieved), then we assumed that the student has successfully answered the test and the testing system calculates a base mark by interpolation using formula of equation of line with the next parameters:

x1 - compulsory level of knowledge(it is forming as % from x2);

x2 - maximal number of points in the test(it is a sum of maximal marks);

y2 - maximal mark in the test by chosen scale;

y1 - minimal mark in the test by chosen scale;

a, b - coefficients in the formula of equation of line;

x - sum of all marks in the range of marks between 3.00 and 6.00;

O1(x) - calculated base mark of the whole test;

STAGE 2
Reduction of the calculated base mark of the whole test
("the real mark")

If in the whole test there are no questions with marks less 3.00 then the final mark is the same as the calculated base mark. In the opposite case the calculated base mark O1(x) is reduced to mark O2(x) using formula of equation of line as reduction equation with the next parameters:

x1 - minimal sum of marks in range < 3.00;

x2 - maximal sum of marks in range < 3.00;

y2 - minimal mark of the test;

y1 - calculated base mark O1(x) by STAGE 1;

a, b - coefficients into the formula of reduction equation of line;

x - sum of all marks in the test <3.00;

O2(x) - calculated (reduced) final mark of the test;

The next three figures illustrate the methods described above and scales:

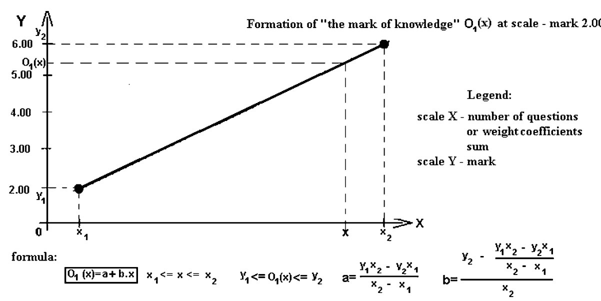


figure 2

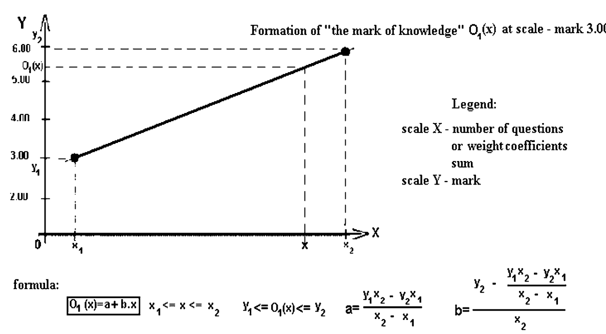


figure 3

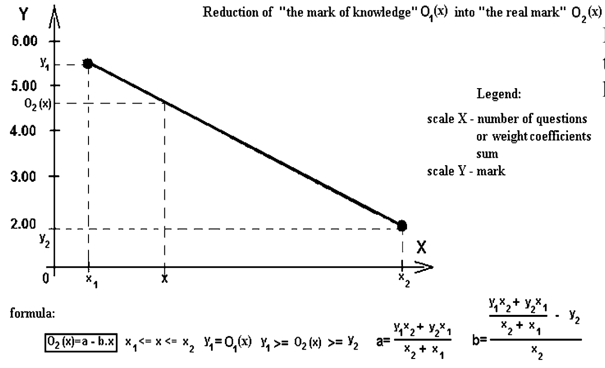


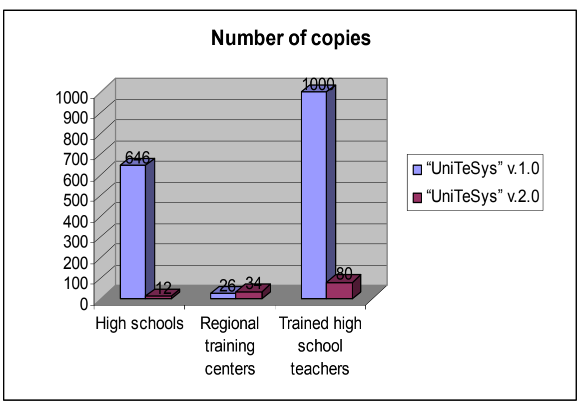
figure 4

V. DISTRIBUTION

The official distributor “Europartner2007” foundation certificated ISO EN 9001:2000 is a specialized primary and high school teacher training center. The distribution of the testing system “UniTeSys” has successfully started since October 2006 and up to this moment the system was developed in two versions. Practical training with the testing system passed over 1000 high school teachers. There are more than 700 installations (copies) at high schools and regional training centers in Republic of Bulgaria (see Table 2). The testing system “UniTeSys” was distributed at 52 high schools and teacher training centers. Effective September 2008 started the distribution of version 2.0.

Table 2
Installations of the testing system “UniTeSys”

Installations	**“UniTeSys” v.1.0**	**“UniTeSys” v.2.0**
High schools	646	12
Regional training centers	26	34
Trained high school teachers	More 1000	more 80



For the best acceptance of the testing system by the teachers there were prepared several video lessons at WWW-http://unitesys.hit.bg (Bulgarian language only)

Table 3
Video lessons

Chapter	Video file, AVI
Installation and Activation of the testing system "UniTeSys"	lesson1.avi
Load Databases of the testing system "UniTeSys”	lesson2.avi
Set weight coefficients of the answers	lesson3.avi
Attach pictures to the questions of the test	lesson4.avi
Transfer loaded test and student tests	lesson5.avi
Run a test	lesson6.avi
Check up a student test	lesson7.avi
Load a list of students	lesson8.avi
Settings of the testing system "UniTeSys"	lesson9.avi
Create an archive	lesson10.avi
Load an archive	lesson11.avi
Statistics	lesson12.avi

VI. CONCLUSION

A test system with predetermined qualities was created. The proposed structure of the system allows reaching given objectives. From October 2006 until now the system is improved in two versions. There are more than 700 plants (copies). The system was distributed in 52 primary, high schools and teacher training centers in Bulgaria. Clean and friendly user interface made easy teacher training with the system. Practical training with the testing system passed over 1000 high school teachers. Using the test system in the practice of the teachers facilitate their work and improve the quality of assessing the knowledge of students. Found increased interest of students in computer test form of assessment. The testing system “UniTeSys” is a suitable and flexible assessing tool. It gives the opportunity to carry out a group test in a computer lab with up to 10 machines in different disciplines. A friendly interface makes the system suitable for primary and high school students at the age of 10 years and above. Built in small statistical module gives the teacher possibility of doing analysis on:

- What is the average mark for each separate question in the test?
- What is the average test mark for the whole student group (class)?
- What is the testing time for each separate question?

Using the system by the customers doesn’t require any special hardware and software resources. This tool gives the teachers the opportunity to assess students fast and easy. Simple for using, for both: students and teachers at different ages.

Improvement on the test system in the next version will be aimed at improving statistical analysis module for evaluation of the test and insert a new module based on artificial intelligence (AI) at Prolog.

REFERENCES

[1] R. S. Nickerson, "New Directions in Educational Assessment", Educational Researcher, December 1, 1989; 18(9): 3 - 7
[2] R. L. Linn, E. L. Baker, and S. B. Dunbar, "Complex, Performance-Based Assessment: Expectations and Validation Criteria", Educational Researcher, November 1, 1991; 20(8): 15 - 21.
[3] John R. *Frederiksen*, "A Systems Approach to Educational Testing", Educational Researcher, Vol. 18, No. 9, 27-32 (1989)
[4] V. Petrov, T. Todorov, "Base of the Statistics", Abagar, Sofia, 2000
[5] I. Syikova, S. Todorova , "Statistical research", Sofia, 2000
[6] G. Bijkov., "Theory and methods in the didactics tests", Sofia, 1996

On the Percolation Behavior of the Thin Films of the PEDT/PSS Complex: a Mesoscale Simulation Study

Toomas Kaevand[a], Jaan Kalda[b], Andres Öpik[a], Ülo Lille[a1]

[a]*Department of Material Sciences, Tallinn University of Technology, Ehitajate tee 5, 19086 Tallinn, Estonia*

[b]*Institute of Cybernetics, Tallinn University of Technology, Akadeemia tee 21, 12618 Tallinn, Estonia*

I. Introduction

The broad range of the electrical conductivities is the characteristic feature of the electronically conducting polymers [1]. Thus, the conductivity of poly(ethylenedioxythiophene)/poly(styrene sulfonate) (PEDT/PSS, Fig. 1) film varies ca over 5 orders of magnitude, depending on the PEDT/PSS mass ratio in the range of 1:2.5 to 1:20 [2,3].The conductivity is attributed to the presence of a conducting, i.e. percolating network, and the film is imagined as a percolating conductor-insulator composite [4–7]. Due to the polaronic nature of the charge carriers in the film [8], this is a rough approximation. However, as shown below, it serves well as a reasonable starting point for the numerical modeling.

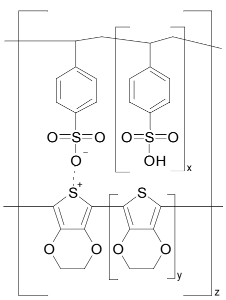


Fig 1. Simplified chemical structure of PEDT/PSS complex.

To our best knowledge, no studies have been previously focused on the percolating pathways in such films. Our mesoscale simulations of this complex are aimed to get a better understanding of this percolating network.

In our first report on the mesoscale simulation of the morphology of the PEDT/PSS complex, the behavior of this complex has been simulated as a polyelectrolyte/ionomer [9] (and references therein). The dynamic density functional theory implemented in MesoDyn simulation code (Accelrys Inc.) has been used. In these coarse-grained simulations (see Fig. 2; the adjusted parameter values are derived from the data of the full atomistic simulation), the morphology evolves according to Langevin stochastic diffusion equations through the minimization of the free energy (periodic boundary conditions are used). In such a way, 3-dimensional density fields ρ(**r**) are obtained. The length and time scales of a mesoscale simulation are up to several hundreds of nm and 10^{-3} s; these numbers match well the parameters of the physical methods used in the morphology research of thin films, and the relaxation times of the polymers.

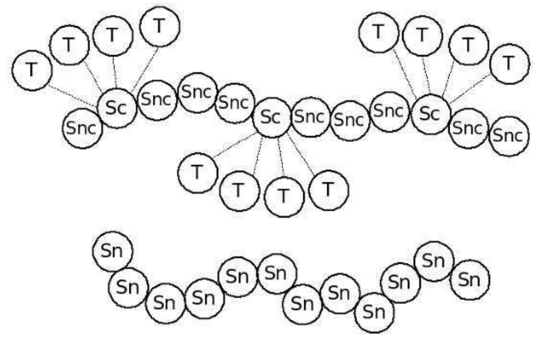


Fig. 2 Upper sketch: T beads interacting with the Hc mesomolecule. Lower sketch: Hn mesomolecule (which are missing in the water-washed films). The beads are marked as follows: T – PEDT, Sc – PSS, Snc and Sn – PSSA .

Further, we used the density fields of the T beads to design a methodology for the evaluation of possible percolating pathways in the thin film of this associate [10]. This methodology enabled us to identify the correlated percolation pathways in the pristine and water-washed films, reflecting qualitatively the differences in the electrical conductivities of these films.

In this study we generalize our calculations of the percolation pathways to a broad range of the PEDT/PSS mass ratios (further called simply "the mass ratio"), doping levels and interaction parameters to get a necessary insight into the percolation behavior of this complex in the framework of the generated morphologies. This allows us also to pinpoint the limitations of the current approach.

II. Generation of the percolation pathways

The generation of the percolation pathways is described in detail in [10] where different topologies of the PEDT-carrying PSS chain have been compared. In this report we use the statistical (random) copolymer model.

For an overview of the percolation theory see [11].

The density fields are generated in a 3×82.3 nm cubic periodic box. In order to integrate the diffusion equation, the

1) Corresponding author, lille@chemnet.ee

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_20, © Springer Science+Business Media B.V. 2010

space is discretized: the box is divided into 32^3 = 32770 cells. All the discrete beads are distributed over these cells, forming thus the density fields of the respective beads. In the volume rendering conditions, these fields appear as the dimensionless volume fraction fields $\eta = \nu\rho$, where ν is the bead volume. Conductivity depends mainly on the distribution of the fields of the T beads. The density fields of the T cells vary within a broad range, from the maximal value d_{max}, down to zero; the conductivities vary in the respective way. This corresponds to a wide range of the barrier heights of the quantum mechanical tunneling process, which is the basic conduction mechanism in the PEDT/PSS complex [12]. Those T cells, which have at least one side in common, form percolation clusters. A percolation pathway is formed by a large cluster spanning from one side of the cube to the opposite side. In the case of an infinite sample, a percolating (infinite) cluster would appear at the critical volume fraction p_c of the T cells, i.e. at the percolation threshold. Our simulated samples are finite; so, the obtained values of p_c are slightly inaccurate. We found such a value Δd (henceforth called "*the conduction window density width*") that the "highly conducting region" with $d_{max} \geq d \geq d_{max} - \Delta d$ forms a percolation pathway. Then, the value of p_c is determined as the volume fraction of that "highly conducting region". Recall that the conductivity range within the sample is very broad; so, one is justified to interpret the lower bound of the conductivity window (i.e. corresponding to $d_{max} - \Delta d$) as the bottleneck conductivity (following the approach of [13]). An infinite cluster consists of the current carrying backbone and the dead ends. The ratio of the numbers of the cells in the infinite cluster and in the backbone to the total number of the cells are referred to as the density of the cluster $P(p)$ and the density of the backbone $P(p)_b$.

In this report, all the morphologies are generated at 298K.

III. Results and Discussion

In Tab. 1, the values of some percolation parameters are shown for different mass ratios, alongside with the volume fractions Θ_T (vol.%, an input parameter) of the T beads, and the corresponding mass ratio values. The conduction window average density Δd decreases together with the mass ratio, and absolute value of its width fluctuates around the average value 0.152 (± 0.039). This is a very natural result, since the hopping mechanism of the conductivity presumes the dependence on the inter-particle distance. Qualitatively, higher bead densities inside the percolating pathway imply higher values of the conductivity (keeping other parameters constant; the significance of the thickness of the isolating Sc layer is ignored). However, there is no direct way of deducing the conductivity from the percolation threshold p_c.

The p_c and $P(p)$ values increase ca three-fold, if the mass ratio changes from 1/1.4 to 1/20. It demonstrates that in the latter case, for the formation of the percolating pathway, much more cells are needed; the infinite cluster consists of a higher number of the cells, as well. The comparison of the $P(p)$ and $P(p)_b$ values shows that the infinite cluster formed at higher p_c values contains more dead ends because the $P(p)_b$ values fluctuate around 8.32×10^{-3} ($\pm1.75\times10^{-3}$). It is clearly seen in Fig. 3 by comparing percolating pathways of items 1 and 4; the corresponding conduction windows and density histograms are shown in Fig. 4, 5. Interestingly enough, the domain size seems to have a maximum at mass ratios 1/2.5 and especially at 1/6. In parallel to this the conducting window width increased as well. It demonstrates the complex relationship between the morphology and mass ratio. The value 1/20 seems to be the lowest one at which the percolation pathway still does exist. Note that the percolating pathways are the fractal objects, which have a non-integer dimension f < 3 (in 3D space) [14].

TABEL I

RESULTS OF CALCULATIONS: THE EFFECT OF THE PEDT/PSS MASS RATIO(DOPING LEVEL 0.25, 298 K, σ VALUES ARE SHOWN FOR THE COMPARISON)

Item	PEDT/PSS mass ratio			
	1/1.4 (w-washed)	1/2.5 (pristine)	1/6	1/20
Δd [1)]	0.812÷0.955	0.648÷0.841	0.340÷0.531	0.050÷0.132
$\lvert\Delta d\rvert$ / Δd	0.143/0.884	0.193/0.745	0.191/0.436	0.082/0.091
p_c [2)]	0.0632	0.0802	0.113	0.199
$P(p)$ [3)] $\times 10^{-2}$	1.51	2.10	2.68	5.20
$P(p)_b$ [4)] $\times 10^{-3}$	4.67	8.12	12.7	7.78
Θ_T [5)]	0.495	0.333	0.143	0.05
σ [6)]	0.15 – 3.0	0.1 – 2.0	5 x 10^{-3}	6 x 10^{-5}

1) the conduction window density width, 2) the percolation threshold, 3) the density of the critical cluster, 4) the density of the backbone (current-carrying core), 5) the volume fraction of the T beads in the system, 6) the electrical conductivity of the complex, S cm^{-1} (according to [2]).

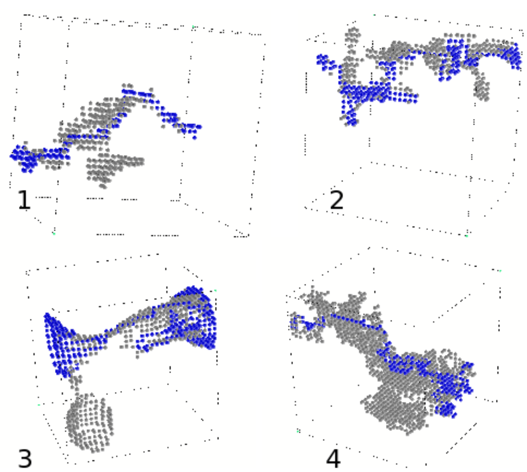


Fig. 3. The percolating pathways (gray dots) and charge carrying backbones (black dots) in the PEDT/PSS thin films. 1,2: the water-washed and the pristine film at the initial PEDT/PSS mass ratio 1/1.4 and 1/2.5 respectively, 3,4: the films at the mass ratio 1/6 and 1/20 respectively.

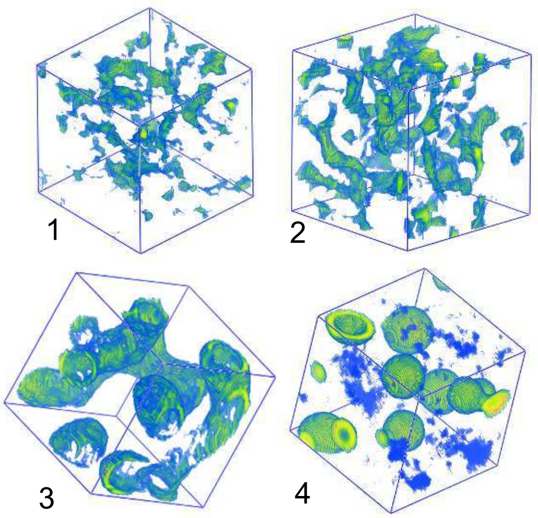


Fig. 4. "Highly conducting" volume renderings of the density fields of the T beads at the mass ratios 1/1.4 (1), 1/2.5 (2), 1/6 (3) and 1/20 (4).

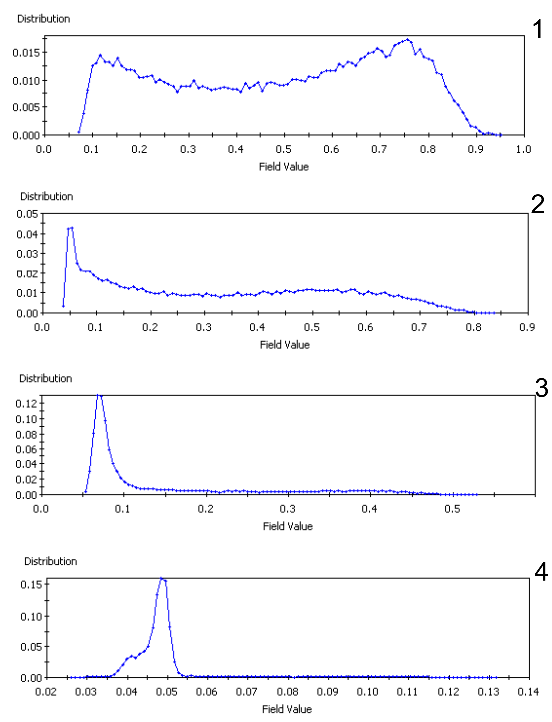


Fig. 5. The density distribution curves of the mass ratios 1/1.4(1), 1/2.5(2), 1/6(3), 1/24(4).

Note that the p_c values do depend on the concentration Θ_T, unlike in the case of the classical percolation problem. This feature is related to long-range density correlations; so, we deal with the correlated percolation problem. We found the site occupation correlation function exponent H values to be in the range of -0.34 ÷ -0.44; c.f. [15]. So, the T cells are "filled" not independently, there are the long-range correlations between them.

The behavior of the $P(p)$ curve near to the percolation threshold is governed by the power law

$$P(p) \propto (p - p_c)^{\beta}\text{ , } if\ p\text{-}p_c << 1. \quad (1)$$

In the case of uncorrelated 3D percolation problem, the exponent $\beta = 0.41$ [11]. Due to the fact that $\beta < 1$, the growth rate of the $P(p)$-curve for $p \approx p_c$ is almost infinite. This is opposite to what has been observed in our case: there is a moderate, or almost a vanishing initial growth rate of the $P(p)$-curve, see Fig. 6. Such a behavior corresponds to $\beta > 1$. The departure from the value $\beta = 0.41$ can be again related to the long-range correlations. In the region where p exceeds considerably p_c, the cluster density $P(p)$ grows nearly proportionally to p (note that the similarity of this plot to that of the conductivity vs. p plot of a random conductor network [11] is just a coincidence). Calculations show that with growing p, the contributions of the densities of the dead ends and of the backbone to the growth of the infinite cluster density are almost proportional (the coefficient of proportionality is 1.06, R^2 0.997). One can speculate that in the regions corresponding to such large values of p, the field density values are lower and hence, these cells do not make a significant contribution to the global conductivity.

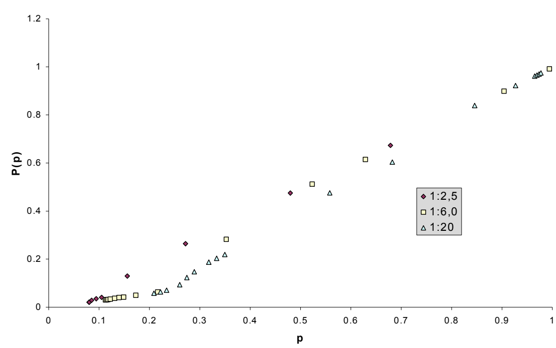


Fig. 6. $P(p)$ vs. p plot at different mass ratios.

In the case of random resistor networks with high conductivity variance, the global conductivity is related to the bottleneck conductivity (the conductivity corresponding to the density level d_{max}-Δd). This value is controlled by the parameter Θ_T. Empirically, we observed a power law: the macroscopic conductivity is proportional to Θ_T^{μ}, with $\mu = 3.70$. Further studies are needed to derive such a dependence from the basic principles.

The dependence of the percolation parameters on the doping level is shown in Tab. 2. Note that the doping level is here defined as the number of Sc beads per one T bead, i.e. the change of the doping level leads to the change in the topology of the Hc mesomolecule (see Fig. 2. The topology enters in various ways into the free energy functional $F[\rho]$ [16]. This results in the shift of the conducting window to slightly lower values of the density (for a constant value of Θ_T). However, the density histogram is dominated now by a more narrow range of densities, see Fig. 7. This is caused by an increasing crowdedness of the T beads on the Hc mesomolecule.

TABEL II

THE RESULTS OF THE CALCULATIONS: THE EFFECT OF THE DOPING LEVEL (PEDT/PSS MASS RATIO 1 : 2.5)[1)]

Item	Doping level			
	0.34	0.5	0.64	0.75
Δd	0.578÷0.707	0.523÷0.643	0.499÷0.580	0.489÷0.557
p_c	0.121	0.116	0.0955	0.123
$P(p) \times 10^{-2}$	2.87	3.36	1.23	1.71
$P(p)_b \times 10^{-3}$	12.6	9.2	5.92	5.8

1) See also Tab. 1.

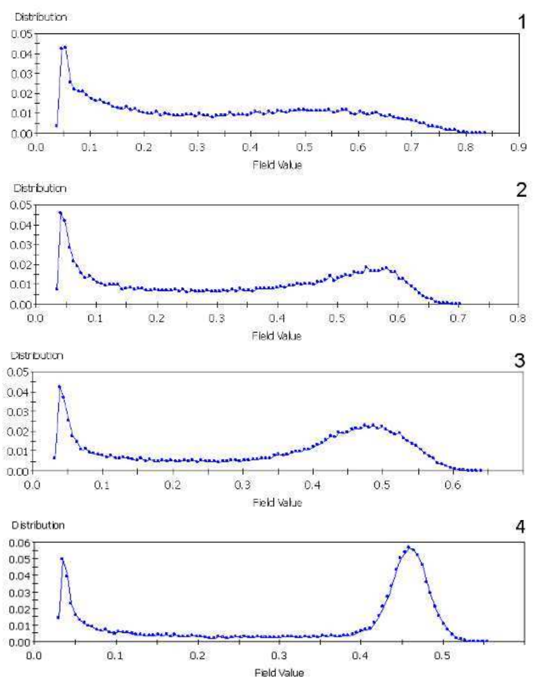


Fig. 7. The density distribution curves 1, 2, 3, 4 at the doping levels 0.25, 0.34, 0.50 and 0.75 respectively.

The percolation parameters are only slightly affected by the doping level (except $P(p)_b$ values). In practical applications, the doping level values 0.25 to 0.33 are typically used; these are expected to be optimal, i.e. above these values the saturation level is reached. (in an polyacetylene this value is about 0.1 [8]) Interestingly, we observe the minimum efficiency of the backbone at the doping levels 0.34 – 0.5.

The doping introduces the charge carriers into the electronic structure, and this problem is beyond the reach of the current approach.

Note that with respect to the interaction parameter values, certain insensitivity is observed. Throughout this report, we have used χRT values −1.4 for the T-Sc electrostatic attraction, and +5.0 for the repulsion of all the other beads. In some experiments, the attraction has been varied up to −5, without detectable qualitative changes. Further increase up to −12 leads to $P(p) = 0.0337$, i.e. nearly the same result as in the case of the standard interaction parameter values; however, there is a considerable growth of the p_c and $P(p)_b$ values, by a factor of 2 and 4, respectively. Note that the simulation quality has been estimated by the smoothness of the free energy minimization curve (it should be without any kinks in the observed interval).

In order to analyze the conductivity enhancement via modifications and annealing [17–20], our approach needs modifications. Our preliminary full atomistic simulations show that one of the possible mechanisms of the solvent effects is the screening mechanism.

In the light of recent refinement of the generally accepted grainy morphology of the PEDT/PSS complex [21] of a specific interest is the treatment of the temperature-dependent percolation behavior in the lateral (parallel to the surface of the substrate) and perpendicular directions separately. By definition in this report we have demonstrated the presence of percolating pathways in the lateral direction, i. e. in the plane of the film in which the conductivity is the highest one. By the way, in our previous simulation study a highly oriented morphology has been demonstrated in the simulations where PEDT-carrying PSS chain has been modeled as an alternating copolymer. (see fig. 5 in [9]).

Finally, the calculation of the absolute values of the conductivity is beyond the reach of the current approach, and remains as the general challenge for further studies.

In this report the thin film of the PEDT/PSS complex has been modeled as a correlated random network of conductors with a large conductivity variance. The application of the percolation concept allows us to elucidate the dependence of the electrical conductivity on the morphology in a broad range of the PEDT/PSS mass ratios at a constant doping level.

CONCLUSION

The morphology of the thin film of the PEDT/PSS complex generated by the MesoDyn simulation code has been treated in the framework of the percolation concept. The main parameters of the correlated conducting network have been calculated for the range of the PEDT/PSS mass ratio from 1/1.4 to 1/20; the effect of the doping level has been evaluated as well. Estimating the percolation parameters proved to be useful and allowed us to obtain new insights into the percolation behavior of the PEDT/PSS complex.

ACKNOWLEDGMENT

The authors are grateful to the Estonian Science Foundation for financial support (Grants No 6633 and 6121) and to the Accelrys Support Group for customizing computational programs.

Reference List

1 A. B.Kaiser , Systematic conductivity behavior in conducting polymers: effects of heterogeneous disorder. *Adv. Mater.* **13**, pp. 927-941 (2001).
2. R. R. Smith, A. P. Smith, J. T. Stricker, B. E. Taylor, and M. F. Durstock, Layer-by-layer assembly of poly(ethylenedioxythiophene): poly(styrenesulfonate). *Macromolecules* **39**, pp. 6071-6074 (2006).
3. S. Kirchmeyer and K. Reuter, Scientific importance, properties and growing applications of poly(3,4-ethylenedioxythiophene). *Journal of Materials Chemistry* **15**, pp. 2077-2088 (2005).
4. S. Ghosh and O. Inganas, Self-assembly of a conducting polymer nanostructure by physical crosslinking: applications to conducting blends and modified electrodes. *Synthetic Metals* **101**, pp. 413-416 (1999).
5. M. Kemerink, S. Timpanaro, M. M. deKok, E. A. Meulemkamp, and M. J. Touwslager, Three-dimensional inhomogeneities in PEDOT: PSS films. *J. Phys. Chem. B* **108**, pp. 18820-18825 (2004).
6. D. M. DeLongchamp, B. D. Vogt, C. M. Brooks, K. Kano, J. Obrzut, C. A. Richter, O. A. Kirillov, and E. K. Lin, Influence of a water rinse on the structure and properties of poly(3,4-ethylene dioxythiophene):poly(styrene solfonate) films. *Langmuir* **21**, pp. 11480-11483 (2005).
7. S. Timpanaro, M. Kemerink, F. J. Touwslager, M. M. De Kok, and S. Schrader, Morphology and conductivity of PEDOT/PSS films studied by scanning-tunneling microscopy. *Chemical Physics Letters* **394**, pp. 339-343 (2004).
8. A. J. Heeger, Semiconducting and metallic polymers: the fourth generation of polymric materials. *Synthetic Metals* **125**, pp. 23-42 (2002).
9. T.Kaevand, A.Öpik, and Ü.Lille. A mesoscale simulation of the morphology of the PEDT/PSS complex in the water dispersion and thin film: the use of the MesoDyn simulation code. Sobh T. Advances in Computer and Information Science and Engineering[XVIII], 540-546. 2008. Springer. CISSE Book Proceedings.
10. T. Kaevand, J. Kalda, A. Öpik, and Ü. Lille, Correlated percolating networks in the thin film of polymeric PEDT/PSS complex as revealed by the mesoscale simulation.*unpublished*.
11. D. Stauffer and A. Aharony. Introduction to percolation theory. Second Edition. 1994. Taylor&Francis.
Ref Type: Serial (Book,Monograph)
12. J. Y. Kim, J. H. Jung, D. E. Lee, and J. Joo, Enhancement of electrical conductivity of poly(3,4-ethylenedioxythiophene)/poly(4-styrenesulfonate) by a change of solvents. *Synthetic Metals* **126**, pp. 311-316 (2002).
13. V. Ambegaokar, P. I. Halperin, and J. S. Langer, Hopping conductivity in disordered systems. *Physical Rev. B* **4**, pp. 2612-2620 (1971).
14. R. Zallen, *The physics of amorphous solids*, John Wiley Sons, Inc., New York, Chicester, Weinheim, Brisbane, Singapore, Toronto (2004), pp.135-191.
15. M. B. Isichenko, Percolation, statistical topography, and transport in random media. *Rev. Mod. Phys.* **64**, pp. 961-1033 (1992).
16. Fraaije J.G.E.M., van Vlimmeren B.A.C., Maurits N.M., Postma M., Evers O.A., Hoffmann C., Altevogt P., and Goldbeck-Wood G.. The dynamic mean-field density functinal method and its application to the mesoscopic dynamics of quenched copolymer melts. J.Chem.Phys. 106[10], 4260-4269. 1997.
17. S. K. M. Jonsson, J. Birgerson, X. Crispin, G. Greczynski, W. Osikowicz, A. W. Denier van der Gon, W. R. Salaneck, and M. Fahlman, The effects of solvents on the morphology and sheet resistance in poly(3,4-ethylenedioxythiophene)-polystyrenesulfonic acid (PEDOT-PSS) films. *Synthetic Metals* **139**, pp. 1-10 (2003).
18. W. H. Kim, G. P. Kushto, H. Kim, and Z. H. Kafafi, Effect of annealing on the electrical properties and morphology of a conducting polymer used as an anode in organic light-emitting devices. *J. Polymer Sci. B* **41**, pp. 2522-2528 (2003).
19. M. M. De Kok and et al, Modification of PEDOT:PSS as a hole injection layer in polymer LEDs. In *Physics of organic semiconductors* ed. by W.Brütting, pp 451-473. WILEY-VCH, (2005).
20. J. Ouyang, Q. Xu, C. Chu, Y. Yang, G. Li, and J. Shinar, On the mechanism of conductivity enhancement in poly(3,4-ethylenedioxythiophene):poly(styrene sulfonate) film through solvent treament. *Polymer* **45**, pp. 8443-8450 (2004).
21. A. M. Nardes, M. Kemerink, R. Janssen, J.A. Bastiaansen, N.M.Kiggen, B.M. Langeveld, J.J. van Breemen, M.M. deKok, Microscopic understanding of the unisotropic conductivity of PEDOT:PSS thin films. Adv. Mater. 19, pp.1196-1200 (2007).

Online Learning Barriers

Jozef Simuth
Vysoka skola manazmentu/ City university of Seattle
Panónska cesta 17, 85104 Bratislava, Slovakia
Ivan Sarmany-Schuller
Slovak Academy of Science
Dúbravská cesta 9, 813 64 Bratislava, Slovakia

Abstract: **The impact of cognitive styles preference and of achievement motivation on the intensity of selected barrier perception in on-line learning was tested on 234 university level student sample. The results of the research don't indicate statistically significant correlations between the achievement motivation and on-line learning barriers perception. The results also show very low correlation between the analytic-intuitive and category width cognitive styles and on-line learning barrier perceptions. The results indicate that students perceive limited opportunities to communicate face to face with classmates and the teacher as a barrier to on-line education. Students also perceived the typically text based study materials and on-line activities as potential barriers to on-line learning.**

I.INTRODUCTION

Despite of the spread of online courses at universities, there is a continuing lack of research in the psychological aspects of online learning. This fact inspired us to carry out research on the pedagogical and psychological aspects of online education. We focused our attention on the possible impacts of cognitive styles and of achievement motivation on selected barriers perception in online learning. We selected two dimensions of cognitive style (analytic-intuitive, category width) which entail social, information perception and acquisition elements. All three characteristics are used as predictors of academic success. Based on the literature review and personal empirical experience we identified five categories of potential online learning barriers: technology, online communication with teacher, online communication with peers, electronic text based study materials and online learning activities.

We formulated research hypotheses focused on finding the relation between the intensity of barrier perception in online learning and two dimensions of cognitive style - Category width and analytic-intuitive dimension of cognitive style. Category width is characterized by broad and narrow categorization. Broad categorizers prefer greater independence, freedom and variety of experience. Narrow categorizers are more successful at tasks which require detailed or analytical skills [2]. Analytic-intuitive cognitive style is measured on two dimensions. Analytic preference is typical for precision, high performance, discipline, conformity and focus on detail. Intuitive preference is characterized by preference for social interactions and cooperation, lack of discipline and tacit approach to problem solving [3], [4].

Next group of hypotheses focused on the impact of achievement motivation on the barrier perception. Furthermore, we formulated research questions that were aimed at identifying the intensity of barrier perception and possible differences between the sample groups.

II.METHOD

We hypothesized that intuitive students would perceive the limited opportunities for creating real social contacts in online learning as barrier more than their analytic counterparts. Moreover, we assumed that intuitive students would perceive the structured text based materials as barrier more than analytic students. Similarly, we presupposed that broad categorizers would perceive the communication and didactical limitations of online education more intensively than narrow categorizers. The achievement motivation characteristic was assumed to have a negative relation with the barrier perception intensity.

We created an Online Barriers Questionnaire (OLB). The OLB is an 18-item questionnaire measuring the intensity of particular barrier perception on the level from Level 1 (no barrier), Level 2 (mild barrier) to Level 3 (extreme barrier that prevents the student from further participation in the course.) The items were divided into 5 categories (table 1).

TABLE I
Categories of online learning barriers

Category	Barrier description	Barrier description	Barrier description
Technology	Use of keyboard	Use ICT as the only tool for learning	Reading from the screen
Communication with instructor	Scarce communication with the teacher (less than twice a week)	Delayed feedback from the teacher	Communication channels limited to email and discussion forum
Communication with classmates	Limited possibilities to communicate with peers	Limited opportunity to compare the knowledge with peers	Communication channels limited to email and discussion forum
Study activities	Limited variability of study activities	Ambiguous instruction	Frequent interruption in studying
Study materials	Only text based information	Rigidity and amount of theoretical information	

We distributed OLB together with the Slovak versions of the Achievement Motivation Questionnaire (DMV) [1]; the Category Width Questionnaire (CW) [2]; Cognitive style index (CSI) [4]. (All in Slovak version). The questionnaires were distributed by email or personally. The anonymity of participants was protected.

We distributed the set of questionnaires to 425 Slovak university students. The return rate was 55%. That means that the research sample counted 234 university students (155

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_21, © Springer Science+Business Media B.V. 2010

women and 119 men, the average age was 22.7). 175 students were students in face-to-face format (AM=21.59, SD=0.13) and 59 students studied online (AM=24, SD=0.23).

All students were informed about online learning process in detail before we administered the questionnaires. The collected data was statistically analyzed. We used Pearson's correlation coefficient r to find the relationship between two quantitative parameters. We used regression analysis for checking the influence of random variables. Arithmetic mean and group median were used for description of barrier perception intensity. Standard deviation was used for revealing the inconsistency of variables.

III. RESULTS

Based on the statistical analysis, we arrived at the conclusion that the chosen personality characteristics had insignificant impact on barrier perception. These results contradicted the hypotheses. We have found out that the preference of cognitive styles on the category width and analytical-intuitive dimensions were not associated with the intensity of particular barriers perception. Similarly, the level achievement motivation had no impact on barrier perception. Despite the fact that our research did not show a link between the two dimensions of cognitive style (analytic-intuitive and category width), we still think that other dimensions of cognitive style and learning styles may have impact on learning on-line.

Investigating the intensity of particular barrier perception, research results have shown that students do not perceive the technology as a barrier in their learning online (AM=1.287, SD=0.54, n=234). Students perceived the online learning activities (AM=1.82, SD=0.75, n=234) and material presented in electronic format (AM=1.71, SD=0.78, n=234) as mild barriers. The frequent interruption of learning and working on assignments was also perceived as a mild barrier (AM=2.132, SD=0.761, n=234).

Barriers that were related to the social aspect of online learning were perceived as significant obstacles to learning. Delayed feedback was perceived as higher than a mild barrier (AM=2.329, SD=0.822, an=234) by 55.6% of students. The limited frequency of contact with the teacher was perceived as medium barrier (AM=2.034, SD=0.799, n=234).

Students perceived the limited possibilities for communicating with peers as above mild barrier (AM=2.260, SD=0.778, n=234) as well as the limited possibilities for comparing the knowledge with peers (AM=2.137, SD=0.801, n=234). Comparing the intergroup differences, we have found differences in barrier perception between the group of face-to-face and online students. The significant differences were in perceiving the limited possibilities for comparing the knowledge with peers (face-to-face students AM=2.38, SD=0.79, n=175; online students AM=1.91, SD=1.91, n=59). The difference is probably caused by different preferences for social contacts between the groups.

We have also found a difference in the perception of large amount of theoretical information as barrier in learning online (face-to-face format: AM=2.14, SD=0.85, n=175; online students AM=1.72, SD=0.66, n=59). We assume that this difference was caused by greater experience of online students who are prevailingly employed and find it easier to apply the theory to real life situations.

Based on the results, we can say that information communication technology is not the crucial element that influences the quality of online education. Students in our sample perceived the social (limited contact with classmates and the instructor), communication (limitations of communication transmitted by ICT) and didactical (limited variety of study activities and materials) aspects of online education as potential barriers in learning. Because these are important findings for online education we carried out additional research. We approached 80 online students with a questionnaire containing open ended sentences. The analysis of their answers confirmed previous findings. Students perceived passive learning activities, invariable study literature, limited possibilities for communication and contact with teacher and classmates as the key obstacles i.e. disadvantages of online education.

IV. DISCUSSION

The results of the research implied the need for creating set of pedagogical principles applicable for online learning. We have come up with 6 principles of online pedagogy:

- Assure and guarantee frequent and regular contact between the teacher and students as well as among students.
- Develop reciprocity and cooperation among students.
- Provide students with prompt positive feedback.
- Create positive and supportive learning environment.
- Respect the diverse talents and learning styles in creeating the learning activities and materials.
- Provide students with clear and high expectations from the beginning of the course.

These principles or recommendations should enable teachers to create online learning environment that would enhance communication among the participants, group work, productive cooperation among students and consequently creating positive relationships and atmosphere. We suppose this can contribute to enhancement of the online learning process.

It is apparent that online education will be employed even more extensively in the future as either university education or as a mean of lifelong learning courses. The advantages and potential of this form of education are advertised in most of the professional literature focusing mainly at the technology and administrative aspects. However, it seems necessary, and our research has proven this, to focus the attention at the limitations and barriers online learning endures.

V. CONCLUSION

The research results show the necessity to focus attention on the limitations and barriers of online education. The scarcity of research focusing on psychological aspects of online learning is apparent. The research attempted to draw the attention of experts in education and information technologies to these aspects. Better generalization of research results would require a much larger sample of university students. Better validity of results could be gained by including students who dropped the online studies. Further research could also focus on cross-cultural differences. We also see an additional area for research by changing the research methodology and focusing on the impact of other personality characteristics on the online education perception. During the research we came across other areas that also need attention of experts in educations. For example, finding relevant answers to questions like: How do teachers perceive online teaching? Is online teaching suitable for all teachers? Is online education really suitable for all fields of academic studies?

We hope that the research will contribute to the development of online learning based on human needs bringing principles of humanistic pedagogy to online education.

REFERENCES

[1] T. Pardel, L. Maršálová, A. Hrabovská, "*Dotazník motivácie výkonu*. Psychodiagnostické a didaktické testy," Bratislava (1984).

[2] T.F. Pettigrew, "*The measurement and correlates of category width as a cognitive variable*," Journal of Personality, 1958, 26, pp 532-544.

[3] I. Sarmány-Schuller, "*Kognitívne štýl a štýly učenia - možnosti aplikácie v edukáci*i," In: SCHOLA 99. Vzdelávanie vysokoškolských učiteľov. Bratislava 1999.

[4] C.W. Allinson, & J.Hayes, "*The cognitive style index: a measure of intuition-analysis for organizational research*," Journal of Management Studies, 33, 1996, pp 119-135.

Good and Bad in Information Retrieval: Information Literacy and Ethics

Olli Mäkinen[1], Marja Mikola[2], Susanne Holmlund[3]
Seinäjoki University of Applied Sciences, P.O.Box 412, FI-60101 Seinäjoki, FINLAND
Vaasa City Library, P.O.Box 235, FI-65101 Vaasa, FINLAND
The Tritonia Academic Library, P.O.Box 331, FI-65101 Vaasa, FINLAND
[1]email : olli.makinen@seamk.fi
[2]email: marja.mikola@vaasa.fi
[3]email: susanne.holmlund@hanken.fi

Abstract – **In this survey article we chart the ethical aspect of information literacy and seek new ideas to develop its teaching. We also define the concepts of information literacy ethics, de facto ethics and mediation. We outline the problems involved in plagiarism and self-regulation and present the best way to overcome the risks the information-flow by means of criticality and critical evaluation of the sources.**

Keywords – **Information literacy, Teaching, Ethics, Plagiarism, Self-regulation, Evaluating sources, Mediation**

I. INTRODUCTION

Standards for students' awareness of information and IT ethics are defined in the 5th Information Literacy Standard. The standards involve data protection, copyright, plagiarism, netiquette, and research ethics in general. However, they are not the most important ethical issues regarding information. The core of information ethics can, in fact, be found from the 3rd Information Literacy Standard, where it is noted that an information literate student evaluates information and its sources critically and incorporates selected information into his or her knowledge base and value system. [1]

Critical thinking and evaluation of sources are an important part of all academic activities, and hopefully, of living in a society. Such thinking is encouraged by philosophy and it is a prerequisite for all changes and development. It is important in both science and life in general. This sets extensive requirements for all educators.

In this article, we discuss the relation of ethics to information literacy. Information literacy is central to the new media and the use of it, to information retrieval, and the ability to ”read media correctly”. The terms media literacy, and communication and information competence are also commonly used.

In the first section of this article (chapters 2-5) we are going to analyze the ethical background of this problem area. That is why we are discussing phenomena like the ethics of the virtual world, de facto ethics, anonymity, and mediation. We are arguing that people's opinions on the right and wrong in the virtual environment are in contrast to those in the real world. Plagiarism is also treated as some kinds of evil, an example of bad behaviour.

Is it possible to improve this behaviour? The authors argue that information literacy contributes to not only individuals' knowledge and critical attitudes, but also to the functioning of societies and values such as democracy.

In the second part of this study we are going to present in general terms some ways of improving the ethical side of information literacy. The titles of these chapters are self-regulation, edification and critical evaluation of source materials.

Self-regulation is one answer to the problem of plagiarism. Edification, which best suits the content requirements of teaching information literacy, is another essential way of improving the operation of information systems. Edification should, however, be of the kind where students understand the advantages they gain from it. As a result, the students should gain insights or a self-reflective attitude towards problems, so that they would be equipped to act ”ethically correct”.

All the authors of this paper are employees at different academic institutions. The authors (MA Marja Mikola, PhD Olli Mäkinen and MA Susanne Holmlund) both teach and plan information literacy education.

II. THE MORAL AND SOCIAL ISSUES OF INFORMATION LITERACY

Ethics is the study of individuals' moral conduct and its grounds. Moral includes individuals' and society's culturally bound conceptions of what is good and bad and what is right and wrong. For morals to exists, individuals must have internalized these conceptions—whether borrowed, inherited, or independently derived.

Originally the word ethics (*ethos*) had two meanings. The first meaning originally referred to complying with the customary ways in dealings between Athenians, slaves, and foreigners. When accepted manners were followed schematically, the actions were deemed ethically correct.

Ethics also had another meaning. Individuals do not act ethically right until having weighed the meaning and implication of their actions and having arrived at the conclusion that the action is either ethically correct or incorrect. Often ethics is seen as a metalevel of morals—

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_22,

ethics supports morals. Ethics is therefore the theory of morals.

Birgitta Forsman concretizes the difference between ethics and morals using an example that both refer to expertise, one to linguistics, the other to political science. Everyone knows how to use a language, although not all of us are necessarily able to explain our language use by means of grammar. Contrastingly, such task is easy for a linguist. [2]

In the problematic area between information literacy and ethics, the role of the educator becomes a very important one. When we speak of edification or education and when educational institutions increasingly invest to self-directed learning, students should be able to realize the benefits that ethically right actions offer. If students realize this by themselves, the educators have got their message across effectively. Here a reference to the previously presented second meaning of ethics can be made, wherein consideration and self-reflection is required of the actor. In this case, educators and teachers must have characteristics typical in Socratic Maieutics. An educator's role is to ”guide a student to the correct information”.

The connection between actions and functionality can be established using various different philosophical explanation models. According to Jürgen Habermas there is a communicative void in the society. He says that the more people and institutions communicate in a society, the more efficiently democracy works [3

Based on these arguments, it can be said that the relationship between information literacy and ethics can be studied in the light of Habermas' communicativity. Information literacy contributes to not only individuals' knowledge and critical attitudes, but also to the functioning of societies and values such as democracy.

III. Ethics in Virtual Environments: Mediation

How does virtuality differ from the real world? In the virtual world, we are only partially present. It is an alternative environment (world) or consists of an indefinite amount of alternative worlds. The virtual world can also be defined as an escape from reality: individuals use technology to create images that, at least for a while, become substitutes for reality. The philosopher Timo Airaksinen says that human beings loose humanity in the Internet and, therefore, also their gender [4]. According to Airaksinen, the key words of virtuality are alone, safe, and without responsibility—in other words, beyond ethics. This means receding from humanity, if ethicality, acting as an ethical creature, is seen as a part of the definition of being human. In the light of Airaksinen's thoughts the ethical part of teaching information literacy greatly needed.

Mediation is an instrument where messages, discussion, discourse, and behaviour are becoming more and more conceptual and abstract and have an effect on our social being. [5]

Jean Baudrillard claims that mediation accelerates, because of the quick shift to virtuality or hyperreality. He pessimistically describes the effect of communication technologies on the human experience of reality. We do not meet each other face to face any more, but in different virtual environments. This has certainly both alienated us from traditional ethical decisions and at the same time brought about new ways to make choices [6] [7]. According to the idea presented by Introna [8] electronic mediation is inducing a sense of hyperreality into our world and thus vitiating our ethical sense of being.

Our life is changing towards a more abstract direction. Mediation is an obligatory mode that controls our lives and ever-increasing complexity in society (the world), and transforms and changes our life-world.

Mediation has an influence on different everyday social and commercial practices and virtual environments. Distance learning is used to respond to the challenges of life-long learning. In distance learning, virtual learning environments are used to rationalize and standardize education. Virtual learning may cause alienation and estrangement, when the traditional social contacts are missing. But it also has its advantages. [9]

Also anonymity provides a possibility to avoid responsibility. Moral obligations are generally tied to the persona. Appearing anonymously has traditionally been related to those who do not ”stand behind their words”. When online, we can use pseudonyms, be anonymous.

Anonymity has been a standard which Internet users take for granted. Anonymity has been so valuable that it is regarded as a basic right in the Internet. It is also used in some applications in distance learning environments. Anonymity protects privacy, but it also has disadvantages. It is very likely that the possibility to operate anonymously has a negative effect on the individual's behaviour on the Internet. Physical presence and familiarity apparently reduce undesired behaviour, while physical distance and anonymity increases it. This most certainly affects the development of the rules which are generally used on the Internet. Wallace, for instance shows, referring to numerous classic studies that our behaviour differs greatly in such situations [10].

IV. De Facto Ethics

According to pragmatism, values are continuously tested against reality [11]. Rosenthal says that pragmatism does not accept any permanent values, and therefore it is suitable for describing virtual ethics and reality. Usually development in value hierarchies is slow, but sometimes it is surprisingly fast. Pragmatism is also based on scientific optimism [12], where it can be said that pragmatic ethics reflects advances in IT.

A common principle applies on both Internet and information technology in general. As soon as a new technology is introduced, everyone is eager to utilize it as quick as possible. There are no significant attempts to create standards to follow, and instead, solutions created by one vendor will quite soon become "de facto standards". Applications are accepted by the general public without being officially defined and accepted, because the latter would take too long.

Similar de facto practice appears to exist in moral issues concerning the Internet. Moral codes are being shaped and introduced taking only pragmatic issues into account. Actors are everything but professionals. A new moral is being created on the Internet constantly; new virtual communities are being born all the time, and these form rules and practices which depart greatly from commonly accepted ethical codes of societies. From an ethical viewpoint, this phenomenon is interesting—it appears that everything takes place faster on the Internet. Pragmatism is a suitable way for describing the development of the Internet and all de facto practices that are being measured based on their suitability alone.

IV. PLAGIARISM

When speaking of research ethics and unethical actions, plagiarism is often the first problem that comes to mind. It has increased rapidly, and information technology and its development does play a part in this trend. In Sweden cases related to plagiarism increased with 90% from 2002 to 2003. [13]

Plagiarism denotes using material or thoughts produced by someone else without giving credit to the person who originally produced the material. Plagiarism in actual texts is often difficult to define, because science is accumulative. All knowledge is to some extent based on previous research. There are also different levels of plagiarism ranging from mild to coarse. [14]

Transferring information to digital form has accelerated misuse and copying. At the same time, tracing plagiarism has also become easier. Internet is filled with digital information and rewriting it or revising it is relatively easy. By copy-pasting, a person can effortlessly compose a plausible essay in ten minutes, without having to read a single page of source material. According to Fedler, close to 80 % of all students have in one way or another came into contact with cheating. [15]

The best remedy against plagiarism is certainly edification. In practice, more attention should be paid on both research ethics and information literacy education.

The effortless copying and modification of information does not only have an impact on information as such, but it also has ethical consequences. The way people look at source criticism, and producers' and creators' rights, is at a turning point. Even though scientific information is always public, digitalization has turned it into a kind of over-public property, which is easy to steal and then publish in one's own name. At the same time, a shift in attitudes can be found: information is no longer viewed as someone's property, but as general information. This contradicts with the fact that information has become the most important production factor.

The supply of information available in digital form is rapidly increasing. The amount of information doubles every three years, and it is presented in electronic form at some point.

The interesting question is: does the increase of information affect its value or content? Usually a product's value decreases, if its quantities increase.

V. SELF-REGULATION

Self-regulation refers to situations in which actors spontaneously create rules and boundaries for their activities. The purpose of self-regulation is to improve or optimize activities. As a concept, self-regulation is the opposite to legislation: legislation is practiced by a state or communities consisting of several states. [16] What then does self-regulation mean in practice? Let us present one example:

The Danish newspaper Jyllands-Posten published caricatures of the Islamic prophet, Muhammad, on the 30th of September 2005. In Islam, regardless of denomination, there prevails a so called ban on pictures, aniconism. Jyllands-Posten declared that the motives for publishing the caricatures were a reaction against self-censorship, resistance against the constraints on freedom of speech in the Arab world, and in addition, the newspaper wanted to bring attention to the attacks on artists in Holland by Muslim extremists. As we know, the effects of publishing the caricatures were far-reaching. The Finnish and Swedish press, however, practiced self-regulation or self-censorship, as it was expressed in some circles. [17]

Self-regulation in information processing refers to those voluntary actions by which unrestricted activity in e.g. information networks are ensured. In a rapidly developing information environment, everything can not be legislated and there is no reason to do so, but the actors employ self-regulation instead.

There is a great demand for self-regulation. It has repeatedly been presented a prediction in which the Internet collapses, the reason being, not shortcomings of information technology or insufficient teleinformatic resources, but content related problems, such as viruses and spam. It has been said that preventing them is very difficult and that self-regulation and edification are the only weapons against unwanted content.

There are many activities on the Internet that are possible only with the help of self-regulation. The production of open source software requires a community in which the members act freely and altruistically for the benefit of the community. The network can also clog, if it is used too much, as transferring files consumes its capacity.

Usually a group that functions well creates its own practices in due time and the need for interference of supervisors lessen. The group becomes independent. The internal regulation of the Internet also creates an unseen control. Most chat rooms file all comments, which, including the participants' email addresses, are searchable. The operators also have the possibility to close down the chat room or to terminate a troublemaker's contract.

Netiquette is derived from the words network and etiquette. It refers to the norms of conduct on the Internet or behavioural rules of an information network. Netiquette advises what is correct and tactful when communicating by email, in chat rooms, or news groups, and what is right or wrong in advertising or commercialism on the Internet. There is no single netiquette approved by everyone, but several collections of codes regarding behaviour on the Internet exist. Every active chat room, social utility, and role playing game has its own set of rules.

Netiquette is one of the best examples of self-regulation. In a manner of speaking, it brings self-regulation to an individual, to a personal level, on which Immanuel Kant (1724-1804) founded his "universal moral law". Kant made people repeatedly ask themselves what would happen to the world if everyone acted like they do? It is easy to draw the conclusion or make a maximum or a universal moral law on this mind game. If I steal, stealing should be permitted for all and everywhere, if I read someone else's emails without permission, the same right should be given to everyone etc.

VI. CRITICAL EVALUATION OF INFORMATION SOURCES

Criticality is often associated with challenging or asking critical questions: why are we acting or thinking in a certain way, could we not act or think differently? On one hand, a critical thinking process must be able to present alternatives in order to create something new. Criticality can therefore be described as an attitude that does not take the present state for granted, but subjugates it to evaluation, asks for reasons and possible alternatives. On the other hand, criticality can be understood as a kind of meta-cognitive skill, which concerns processing and analysing information, especially when evaluating the fundamentals of prior information.

Criticality requires creativity, conation to know, and autonomy. Creativity is the prerequisite for criticality, independent imagination is required, so that people not only adapt existing views, but see beyond them and find new possibilities. Independence is therefore essential for criticality. A prerequisite for advanced cultures is that people's thoughts are not subdued.

The general aim of pedagogy is to guide towards criticality. In the western society, criticality is a central cultural ideal. Education should develop individuals, so that they become critical towards intellectuals and social systems. Accordingly, the requirements of criticality are primarily defined in the objectives of higher education, where it is emphasized—among other things—that students must understand the basic concepts, information landscape, and methods of their own discipline. They have to learn to analyse, evaluate, and apply information of their own field. They have to become independent, critical, and creative information retrievers and users, who develop themselves continuously. Becoming a critical thinker concerns not only those who are aiming for an academic career. The inevitable consequence of tasks and the constantly changing information related to them is that continuous learning and reflection becomes inevitable in practical professions as well. [18] Critical thinking will continue to be central in working life also in the future—a completely different matter is, whether it is even possible to observe critical thinking in work life, because an employee with independent thoughts and values is a challenge to the work community as a whole. [19]

The concept of criticality as an element of the new (information) literacies is clearly linked to critical thinking skills, and from a communication's perspective, to the tradition of edification. In other words, criticality is connected to information retrieval and critical evaluation of information sources, as Donald J. Leu et al. explain. The fundamental motive is to manage information flows and a clutter of political, ideological, or financial messages from open networks that have to be critically analysed. [20]

David Bawden thinks that critical attitude towards information includes the following five components [21]:

1) Separating the information from facts.

2) Understanding what type of an information source is in question and under what premises it has been produced.

3) Evaluating the usability, topicality, accuracy, and consistency of the information

4) Readiness to continue a query when the first search result is not satisfactory

5) Critical attitude towards results achieved by information retrieval technology.

VII. EDIFICATION

The connection between edification and the information literacy project is self-evident. One could say that the aspect is pragmatic. When individuals and the community act ethically right, the actions are both just (ethically good) and effective (qualitatively good, measurable).

Due to this, information literacy projects are continuously launched, not only in Finland but all over the world. Edification and information literacy education is also important for the citizens from a legal protection standpoint; we see more and more public services online as time advances and especially the elderly and the handicapped no doubt are in danger of ending up in the digital gap. Guaranteeing them service (knowledge, instruments) is an ethical question.

Communality is a virtue in today's society. Teams are becoming increasingly common in work communities. It is considered self-evident that the results of teamwork are more extraordinary than when all employees are working independently. But teamwork is not successful, if no appropriate rules exist. Ethics can be based on satisfying communal needs, and if applied as information literacy, it means taking cognizance of all parties.

Self-regulation, or the ethically correct actions that appear in the information literacy standards, are to a great extent questions of self-reflectivity or self-realization. Understanding and internalizing complex information, regulations, or lingos is easy to no one. For example, copyright laws are constantly changing and are—in the minds of the citizens—open to interpretations (even though it is actually more or less unambiguous). This is where information literacy education comes in.

The preliminary research done in Tritonia shows that after receiving education in ethical issues in information literacy, the students knew that plagiarizing is wrong and that it consists of a complex process where several parties suffer—including the plagiarist [22]). The purpose of the research was to evaluate and develop the information literacy education given at Tritonia.

A sample of 80 students from the first year students at Vaasa University were interviewed about their attitudes towards plagiarism using a before/after method in the autumn 2004. The same questions were posed before and after a lecture on research ethics. The respondents were given the opportunity to express their attitudes toward e.g. file sharing in their own words.

After the lecture, the respondents clearly had a deeper understanding about acting in a network, in a collective environment that is, where unethical actions also affect the situation of other actors. When it comes to plagiarism, also fairness and justice are in question. The students understood that plagiarism damages both commercial quarters (publishers) and information producers (authors, researchers), who all loose something in the process

VIII. CONCLUSION

In this study, we have analyzed and evaluated ethics in connection to information literacy and use of information, including information retrieval and information processing. According the Information Literacy paragraphs, information literate students should be able to critically evaluate information and its sources and be able to incorporate chosen information into their own knowledge base and value system. The paragraphs also involve and regulate data protection, copyright, plagiarism, netiquette, and research ethics in general.

The aim of this article has been to demonstrate how ethical aspects become increasingly important in information environments. Theft, fraud, and plagiarism are threatening the information society's functioning.

Information literacy contributes in many ways to the functioning of our society. Information literacy can, in general terms, be seen as tuning certain processes. From this view point, it also appears to be about effectiveness. It can be argued whether or not efficiency is ethically valuable—it often leads to burn-outs, impersonality, unemployment, and monotonous work. Effectiveness does not go hand in hand with ethical goodness.

Another negative aspect connected with effectiveness and information literacy is mediation. By means of mediation, messages, discussion, discourse, and behaviour are becoming more and more conceptual and abstract, and that affects our social being. As a result of mediation, there is no first-hand experience of reality; everything is constructed, and in virtual reality, we have receded far from real life. Mediation affects our capability to make independent ethical decisions.

In distance learning, for instance, virtual learning environments are used to rationalize and standardize education. Virtual learning may cause alienation and estrangement, when traditional social contacts are missing. We have pointed out in this article that mediation leads to impersonality or "facelessness".

The main focus in this study has been to develop and map different kinds of concepts and actions that exist behind "information literally good ethical conduct". We have interlinked information literacy with moral and ethical issues. The role of educators becomes very important in that area. When edification or education are in question and when educational institutions increasingly invest on self-directed learning, students should be able to realize the benefits that ethically correct actions offer. This is especially important when using information networks. The functioning of a network depends on its users' ability to understand the connection between correct actions and functionality. If students realize this by themselves, the educators have got their message across effectively.

The improvement of information literacy contributes to not only individuals' knowledge and critical attitudes, but also to the functioning of societies and values such as democracy.

As a result, the ethics of the information age differ from traditional ethics. In the third chapter, we introduced the concept of de facto ethics, referring to pragmatism in which values are continuously tested against reality. We have set ethical change against technical development in general. As soon as a new technology is introduced, everyone is eager to utilize it as soon as possible. There are no significant attempts to create standards to follow, and instead, solutions created by one vendor will quite soon become "de facto standards". A similar de facto practice appears to exist in moral issues concerning the Internet. Moral codes are being shaped and introduced taking only pragmatic issues into account. Actors are anything but professionals. A new moral is being created on the Internet constantly; new virtual communities are born all the time, and these form rules and practices which depart greatly from commonly accepted ethical codes of societies.

We also introduced a serious ethical problem, namely plagiarism. As we stated earlier, ethical sensitivity has weakened, when people no longer meet others face to face. Plagiarism and other types of unethical behaviour have become easier. In Sweden, cases related to plagiarism increased 90% between 2002 and 2003. We have also presented ways of preventing a range of unwanted, unethical conduct.

These remedies are criticality and critical evaluation of information sources, self-regulation, and edification. All these activities can be adapted easily for ethics of information literacy. Critical evaluation of information sources is the foundation of all science and is necessary for everyone living in this age of information flow and mass communication. Netiquette is a good example of self-regulation, and edification is what education of information literacy can be called.

We have also outlined a problem which depends on and arises from the fast development in the information environments. Information has become the most important resource in the world. It is now less important who owns the means of production, for instance raw materials, factories, and railways, than it was in the past. The winner is the one who is in possession of or administrates the essential information, for instance patents, workshop drawings, and logistic chains. As a result, misusing information has increased. Survival in the middle of the information flow requires critical thinking and ability to evaluate a variety of information sources. Ethics and information intertwine, as we have highlighted, and all chapters of this article open gates to new, future studies.

REFERENCES

[1] Association of College and Research Libraries (2001). *Information Literacy Competency Standards for Higher Education.* Retrieved August 13th, 2008, from http://www.ala.org/ala/acrl/acrlstandards/informationliteracycompetency.cfm

[2] Forsman, B. (1997). *Forskningsetik. En introduktion.* Lund: Studentlitteratur, pp. 10-11.

[3] Salter, L. (2003). Democracy, New Social Movements, and the Internet: A Habermasian Analysis. In L. L. Brennan & V. E. Johnsson (Eds.), *Cyberactivism: Online Activism in Theory and Practice.* (pp. 117-144). New York: Routledge, pp. 117-118.

[4] Airaksinen, T. (2005). *Ihmiskoneen tulevaisuus.* Helsinki: WSOY.

[5] Mäkinen, O. (2008). Philosophy of Web-Based Meditation. In Hansson, T. (Ed.) *Handbook of Research on Digital Information Technologies: Innovations, Methods, and Ethical Issues.* (pp. 463-479). Hershey: IGI Global.

[6] Baudrillard, J. (1994). *Simulacra and Simulation.* (G. Glaser, Trans.). Ann Arbor: University of Michigan Press, pp. 20-21.

[7] Chandler, D. (1995). *The Act of Writing. A Media Theory Approach.* Aberystwyth: University of Wales, p. 89

[8] Introna, L.D. (2002). The (im)possibility of ethics in the information age. *Information and Organization*, 12, 2, 71-84.

[9] Mäkinen, O. (2008), p. 264.

[10] Wallace, P. (1999). *The Psychology of the Internet.* Cambridge: Cambridge University Press, pp. 239-242.

[11] Rosenthal, S. B. (1999). *Rethinking Business Ethics: A Pragmatic Approach.* Cary: Oxford UP, p.2.

[12] Pihlström, S. (1997). *Tutkiiko tiede todellisuutta? Realismi ja pragmatismi nykyisessä tieteenfilosofiassa.* Helsinki: Helsingin yliopiston filosofian laitos, pp. 11.12.

[13] Report by the Swedish National Agency for Higher Education 2004:17 (Sammanställning av beslut och domar i disciplinärenden som rör studerande vid universitet och högskolor från 2003).

[14] Mäkinen, O. (2007). Informaatiolukutaito ja etiikka. [Information Literacy and Ethics]. In A. Nevgi (Ed.), *Informaatiolukutaito yliopisto-opetuksessa.* (pp. 53–70). Helsinki: Helsingin yliopisto.

[15] Fedler, F. (2006). Plagiarism Persists in News Despite Changing Attitudes. *Newspaper Research Journal*, 27, 2, 24-37.

[16] Mäkinen, O. (2006b). Näkökulmia tietoverkkojen itsesääntelyyn. [Some points of view of the self-regulation in information networks]. *Informaatiotutkimus*, 25, 2, 41-52.

[17] Ibid.

[18] Kautto, V. (2004). *Tieteellisen kirjallisuuden arvioinnin ohjaus yliopisto-opetuksessa: neljän tieteenalan tarkastelu.* [Abstract: Teaching research literature evaluation in higher education: analysis of four disciplines]. Oulu: Oulun yliopisto.

[19] Järvinen – Taubert, J. (1999). Tarvitseeko työelämä kriittisiä työntekijöitä. In J. Järvinen-Taubert, P. Valtonen (Eds.). *Kriittisyyteen kasvu korkeakoulutuksessa.* Tampere: Tampereen yliopisto.

[20] Leu, D. J. Jr., Kinzer, C. K., Coiro, J. L., Cammack, D. W. (2004). *Toward a Theory of New Literacies Emerging From the Internet and Other Information and Communication Technologies.* Retrieved August 25, 2008, from http://www.readingonline.org/newliteracies/lit_index.asp?HREF=leu/index.html

[21] Bawden, D. (2001). Progress in documentation – Information in digital literacies: a review of concepts. *Journal of Documentation*. 57, 2, 218 – 259.

[22] Mäkinen, O; Kuoppala, H. & Rintamäki, K. (2006). Informaation luonne ja plagiointi: empiirinen tutkimus opiskelijoiden käsityksestä tiedosta, sen käytöstä ja heidän suhtautumisestaan virtuaaliopetukseen. In Mäkinen, O. (Ed.) *Akateemisia opiskelutaitoja oppimassa: Opiskelijoiden kokemuksia verkkokurssilta ja tiedonhakutaitojen kehitys lähiopetuksessa.* (pp. 59-118) Vaasa: Vaasan yliopisto. Retrieved August 29th, 2008, from http://www.uwasa.fi/opintoasiat/kehittaminen/w5w-waasa/raportteja/aot

Formative Research for Enhancing Instructional and Design Methods: A Focus on Virtual Reality (VR)-Based Learning Environment

Chwen Jen CHEN
Faculty of Cognitive Sciences and Human Development
Universiti Malaysia Sarawak
94300 Kota Samarahan
Sarawak MALAYSIA
cjchen_usm@yahoo.com or cjchen@fcs.unimas.my

Abstract- This paper reports the process of enhancing the instructional and design methods prescribed in an instructional design model for virtual reality (VR)-based learning environment. The study uses formative research to formatively evaluate the methods with the aim to generate hypotheses on how to improve the robustness of the model. The analysis of the data collected has revealed four principles that are hypothesized to be able to improve the model.

I. INTRODUCTION

Instructional design is a field that concerns with understanding, improving, and applying methods of instruction to bring desired changes in learner knowledge and skills [1]. The goal of instructional design is to produce a well-designed instruction that saves time and money as well as eases the learning process [2]. Indeed, the ultimate aim of any instructional design efforts is to create effective and efficient instruction.

Reference [3] proposes an instructional design model for guiding the design of non-immersive VR-based learning environments. The model prescribes instructional methods to optimize desired instructional outcomes in learning environments that are enabled by the non-immersive VR technology. Fig. 1 shows an illustration of the model that serves as a feasible and useful template to guide the design of VR-based learning environments. The macro-strategy prescribes instructional methods that concerns with the selection, sequence, and organisation of the subject-matter topics that are to be presented and the micro-strategy prescribes design methods concerns with the strategies for effective presentation of the learning contents.

II. FOCUS OF THE STUDY

This study aims to employ formative research to formatively evaluate the instructional and design methods prescribed by model in order to produce hypotheses on how to improve the model. Hence, the specific objectives of this study are to:-

a. improve an instructional design model for VR-based learning environment using the formative research methodology.
b. hypothesize an improved instructional design model for VR-based learning environment based on the outcomes of the formative research.

The overall study focuses on the enhancement of both macro-strategy and micro-strategy. Nevertheless, this paper focuses only on the enhancement of the design methods prescribed by the micro-strategy.

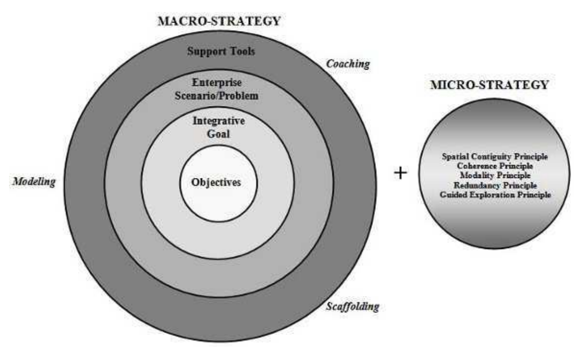


Fig. 1: The instructional design model (Adapted from [3])

III. METHOD

Reigeluth and Frick (1999) proposes formative research as a research methodology for developing as well as improving design theories (or models). Formative research is qualitative and iterative in nature. As pointed out by Richey and Klein (2007), the focus of formative research is to explicitly relate theory/model formation or improvement to the research process. Studies such as English and Reigeluth (1996), Kim (1994), Lee and Reigeluth (2003) have employed formative research to improve instructional theories or processes.

Acknowledgement: The author acknowledges the financial support rendered by UNIMAS through Fundamental Research Grant Scheme FRGS/05(02)/616/2006(49), Ministry of Higher Education, Malaysia.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_23, © Springer Science+Business Media B.V. 2010

This study employs a formative research using a designed case (Reigeluth & Frick, 1999), in which the instructional design model is intentionally instantiated and then formatively evaluates the instantiation. The underlying logic of formative research is that, if an accurate application of an instructional design model is created, then any weaknesses that are found in the application may reflect weaknesses in the model, and any improvements identified for the application may reflect ways to improve the model, at least for some subset of the situations for which the model is intended (Reigeluth & Frick, 1999). This study relies mainly on qualitative data with limited use of quantitative data.

Procedure

Based on Reigeluth and Frick (1999), after selecting a design theory (or model), the formative research process continues by designing an instance of the model. A design instance refers to a specific application of the design model. In this study, the design instance was a VR-based learning environment for novice car drivers to learn about traffic rules and traffic signs. This is followed by the collection of formative data from the representative learners of the design instance via one-to-one evaluation sessions. The process then requires the collected data to be analysed. The intent is to identify and remove problems in the instance, particularly in the methods prescribed by the model as well as to reconfirm the appropriateness of the methods. Revisions made on the instance represent hypotheses as to ways in which the design model itself might be improved. This data collection, analysis, and revision process is repeated several cycles before tentative revisions of the model are proposed.

Formative Data Collection Techniques

The study employed one-to-one evaluation technique to collect formative data. One-to-one evaluation allows probing into the reactions and thinking of the each participant through think-aloud method. During the evaluation, the study incorporated guided interview to identify strengths and weaknesses in the design instance, explore improvements for elements in the design distance, explore consequences of removing elements from, or adding new elements to, the instance, and explore possible situationalities (ways that methods should vary for different situations).

The study also involved each participant by interviewing him or her at the end of the evaluation session. This is also known as debriefing. The collected data was highly insightful and useful or at a minimum, they provided some hypotheses worthy of testing with subsequent participants and situations.

The study also used observation to verify the presence of elements of the design model and to see surface reactions of the participants to the elements. To assist this observation process, the one-to-one evaluation sessions were audio and video recorded using the Noldus Observer system, which was a software package for collection, analysis, presentation and management of observational data.

Having observation, think aloud and guided interview during the one-to-one evaluation session as well as debriefing at the end of the one-to-one evaluation session allow multiple methods of collecting data, which is a form of triangulation. The rationale of having such triangulation is to compensate the flaw of one method with the strength of another.

Instruments

This study involved the use of two sets of instrument. The interview question set guide interactions with participant during the one-to-one evaluation sessions. Debriefing question set provides an opportunity for participants to provide their additional feedback or comments at the end of the evaluation sessions.

Elements of the Design Instance

This study identified three variables of the design instance: navigation speed, environmental richness and collision detection. These three variables were chosen as they were the common features of a three-dimensional virtual environment.

VR allows its user to perform real-time navigation through the three-dimensional virtual environment and this lead to the question on how fast the learner should be allowed to navigate or the preferred speed that will not inhibit the learning process. The study also looked into the effects of the complexity of the virtual environments as well as the system capability to detect collision in the virtual environment.

The researcher manipulated these variables to determine the settings that were perceived by the subjects as user-friendly and/or help (or at least will not interfere) their learning process. The ultimate aim of manipulating such variables is to generate hypotheses that may refine the micro-strategy of the instructional design model.

Variables

A. Navigation Speed

In this regard, the study involved the creation of three learning environments to find out whether the speed differences were noticeable and the preferred speed if they were noticeable. With all other features being similar, the navigation speed for all the five virtual road scenarios available in the learning environments was set to fast, medium and slow respectively.

B. Environmental richness

The study involved the creation of two learning environments. With other features being the same, one of the learning environments incorporated virtual objects, such as trees, buildings, bushes, lake, and etc. while another learning environment did not include all these additional virtual objects. The study is to find out whether such differences are perceptible and the preferred environment if they are perceptible.

C. Collision detection

The investigation into this variable involved the creation of two learning environments. With other features being the same, a learning environment possessed the collision detection feature while another learning environment allowed the learner to move through any virtual objects.

Procedure

For each variable, the study required every participant to explore the first virtual road scenario with the pre-determined settings respectively. For example, in investigating the navigation speed, the subject used fast speed when navigating the first virtual road scenario and then repeated with medium speed as well as slow speed for the same scenario. If the subject was able to notice the differences and decided on his or her preferred setting, the experimenter terminated the use of the learning environments and proceeded to the debriefing session. Otherwise, the study repeated the evaluation with the subsequent virtual road scenario. Each variable involves three participants to reconfirm the findings.

IV. FINDINGS AND INTERPRETATIONS

Findings

This section reports the findings of the one-to-one evaluation sessions and observations of the recorded sessions for a number of variables related to micro-strategy. It then includes the interpretations of these findings.

A. Comments and Observations

The feedback obtained through the guided interview between the participants and the interactant during the one-to-one evaluation revealed consistent responses for elements on navigation speed and collision detection while responses for environmental richness were inconsistent. All the participants of the navigation speed variable preferred the normal speed and all the participants of the collision detection variable were unable to note the difference between the 'on' and 'off' category.

Recording done using the Noldus Observer System had revealed that fast navigation caused difficulty to navigate the car (out of control). Observation on the recording also showed that the presence or absence of collision detection did not seem to be noticeable. When the collision detection was turned on, a participant would not be allowed to navigate out of a road as some transparent blocks were put up to prevent such occurrence. However, when it was turned off, a few participants were observed to navigate along the road side.

B. Debriefing

The responses to the debriefing questions revealed two major issues. Participants raised their concern particularly on the type of input used to navigate and interact with the virtual road scenarios as well as the clarity of the image and text on the virtual traffic signs.

Interpretations

A. Navigation Speed

The finding revealed consistent preference of medium navigation speed for navigating through the virtual environments of the learning environment. As it is unlikely to claim a standard preferred measurement for navigation speed, the consistent preference for medium speed used in the design instance points to the existence of a specific range of speed that is preferable in such navigation activity. Involving a few users in the design and development process of a learning environment will help to determine this appropriate range of speed. Thus, it is hypothesized that a principle of the micro-strategy is that appropriate navigation speed has be provided to enable accurate and easy navigation through the three-dimensional virtual environments.

B. Collision Detection

The finding also revealed that participants were unable to note the difference between virtual environments with collision detection and virtual environments without it. As the virtual environment is simulating a context that is available and observable in the real world, participants did not have the tendency to purposely collide with any virtual objects available in the virtual environment. This could possibly explain their inability to note the existence of the collision detection feature.

The observation of the recording somehow points to an interesting inference. Although collision detection does not seem to be noticeable, depending on the context presented in the learning environment, the collision detection feature can be used as an intervention to avoid undesired misconceptions. In the earlier example, without collision detection, the virtual environment allowed the participant to easily navigate out of the road side. In this context, allowing the participant to navigate her virtual car partially out of a road might cause misconception if the learning environment did not provide any kind of intervention to explain that such navigation was actually against the rule. In such case, placing transparent block along the road sides with the collision detection being turned on, as in the virtual environment with collision detection, would help to prevent the participant from navigating out of the road. In line with this, it is hypothesized that detecting collision is not crucial in a virtual environment that simulates real world context. Nevertheless, detecting collision could sometimes serve as an intervention to eliminate or prevent erroneous understanding while navigating and interacting with a virtual environment.

C. Environmental Richness

The inconsistency in the preference for environmental richness does not help to hypothesize additional specific principles for the instructional design model. These findings indicate the need for further investigations. These may include conducting a survey on a larger sample size to check whether there is a specific mode of preference for each variable and test whether there is correlation between the mode of preference and different situationalities, e.g. individual differences and context differences. Such further investigations may then help to derive additional hypotheses of principles for the micro-strategy.

D. Input Type

The debriefing session had uncovered input type as one of the major arising issues. Many reported difficulty in controlling the mouse for navigating through the virtual environments. This helps to hypothesize the need to have an alternative input type. This study can be further extended to include investigations to determine the most appropriate input type or the need to use additional navigational metaphor to cope with the missing dimension of the two-dimensional input devices used to navigate in a three-dimensional virtual environment.

E. Clarity

The debriefing session also raised the issues on the clarity of the images and text that appeared on the traffic signs used in the virtual road scenarios. Thus, another hypothesized principle of micro-strategy is to incorporate clear images and legible text into virtual environments, particularly if the learning environment emphasizes those elements.

Hypotheses to Improve the Instructional Design Model

The study has revealed four hypotheses related to the principles of the micro-strategy. The navigation speed principle suggests that VR-based learning environment should provide appropriate navigation speed to enable accurate and easy navigation through the three-dimensional virtual environments. The collision detection principle points such detection is not crucial in a virtual environment that simulates real world context. Nevertheless, detecting collision could sometimes serve as an intervention to eliminate or prevent erroneous understanding while navigating and interacting with a virtual environment. The findings on the environment richness variable are immature to lead to any hypothesis. However, debriefing sessions has discovered the need to identify a more appropriate input type than the two-dimensional mouse or to use additional navigational metaphor to cope with the missing dimension of the two-dimensional input devices used to navigate in a three-dimensional virtual environment. It also points to the visual clarity principle that emphasizes the incorporation of clear images and legible text into virtual environments, particularly if the learning environment emphasizes those elements.

V. CONCLUSION

This paper reports the findings obtained via the manipulation of three variables; navigation speed, collision detection, and environmental richness, which are the common features of any three-dimensional virtual environments. It has successfully generated four hypotheses to enhance the micro-strategy of the model. Indeed, further investigations that include other application specific variables as well as the instructional methods prescribed in the macro-strategy are likely to provide more comprehensive insight into the generation of more hypotheses that could further enhance the robustness of the instructional design model.

REFERENCES

[1] G.L. Gropper, "A metatheory of instruction: A framework for analyzing and evaluating instructional theories and models," in *Instructional-Design Theories and Models: An Overview of their Current Status*, C.M. Reigeluth, Ed. New Jersey: Lawrence Erlbaum, 1983, pp.37-53.

[2] G.R. Morrison, S.M. Ross, and J.E. Kemp, *Designing Effective Instruction*. USA: John Wiley & Sons, 2007.

[3] C.J. Chen, S.C. Toh, and M.F. Wan, "The theoretical framework for designing desktop virtual reality-based learning environments," *Journal of Interactive Learning Research*, vol. 15, no. 2, 2004, pp. 147-167.

[4] C.M. Reigeluth and T.W. Frick, "Formative research: A methodology for creating and improving design theories," *Instructional-Design Theories and Models – A New Paradigm of Instructional Theory*, C.M. Reigeluth, Ed. New Jersey: Lawrence Erlbaum, 1999, pp. 633-652.

[5] R.C. Richey and J.D. Klein, *Design and Development Research: Methods, Strategies, and Issues*, Mahwah, NJ: Lawrence Erlbaum, 2007.

[6] R.E. English and C.M. Reigeluth, "Formative research on sequencing instruction with the elaboration theory," *Educational Technology Research & Development*, vol. 44, no. 1, 1996, pp. 23-42.

[7] Y. Kim, *Formative research on the simplifying conditions method for task analysis and sequencing of instructional content*. Unpublished dissertation, Indiana University Graduate School, Bloomington, IN, 1994.

[8] J. Lee and C.M. Reigeluth, "Formative research on the heuristic task analysis process," *Educational Technology Research & Development*, vol. 51, no. 4, 2003, pp. 5-24.

Results on H-Infinity Static Output-Feedback Control of an Electro-Mechanical System

Jyotirmay Gadewadikar[+], Ajay Bhilegaonkar[*], Frank Lewis[**], Ognjen Kuljaca[+]
Systems Research Institute, Department of Advanced Technologies, Alcorn State University
[+]1000 ASU Drive #360, Alcorn State, MS 39096, USA
jyo@alcorn.edu, okuljaca@alcorn.edu
*Pharmaceutical Research Institute, Bristol Myers and Squibb
ajay.bhilegaonkar@gmail.com
[**]Automation & Robotics Research Institute, The University of Texas at Arlington
lewis@uta.edu

Abstract- **This paper presents H-Infinity static-output feedback controller implementation comparison results on an electro-mechanical system test bed. In eletro-mechanical systems it is expensive and sometimes difficult to achieve and implement full state-feedback. Uncertainties in the electromechanical systems and the disturbances affect the performance of the optimal controller; controllers may require prescribed desirable structures as well. In this paper simplified necessary and sufficient conditions are revisited for static output-feedback control for a spring mass damper electromechanical systems. This paper shows implementation of a numerically efficient solution algorithm to solve coupled design equations to determine the H-Infinity output-feedback gain. No initial stabilizing output feedback gain is needed. Effectiveness of the proposed method is demonstrated through comparison of optimal output-feedback, H-Infinity output-feedback and state feedback implementation on spring-mass damper electromechanical system.**

I. INTRODUCTION

H-Infinity control techniques and static output-feedback both are very rigorously research problems in Systems and Controls Community. Static output-feedback controllers are less expensive to implement and more reliable in practice. In particular, static output-feedback controllers can be employed as back up controllers, i.e. as controllers which are not active during the regular operation, but which are used in the case of faults. For this reason back-up controllers have to be as simple as possible [1]. Issues of importance to the design of a practical static output-feedback controller include (i) computational simplicity, (ii) operation solely with measurable quantities, and (iii) implementability in finite time. However, implementation experiences and results on industry like test beds have rarely been published in popular literature. In this paper we offer a design procedure which yields practical and implementable H- Infinity output-feedback controller and addresses above issues.

An excellent survey of static output feedback control is given by Syrmos et al. [20]. It is well known that the OPFB optimal control solution is prescribed in terms of three coupled matrix equations [16]. A sequential numerical algorithm to solve these equations is presented in Moerder and Calise [17]. This algorithm is extended and used for aircraft control design in [19]. An initial output-feedback gain is needed which is difficult to determine for practical systems. Some recent results address these issues; Jen Te Yu [23] provides a solution for the optimal static output-feedback problem in terms of two algebraic Riccati Equations plus a projective coupling condition. An Algorithm is presented that requires an initial stabilizing state feedback gain, which is easy to find using standard Linear Quadratic Regulator (LQR) methods. Yang and Moore [22] give an algorithm for static output feedback based on alternating projections. Output-feedback stabilizability conditions that only require the solution of two coupled matrix equations are given in [21] and [14]. Some recent LMI approaches for OPFB design are presented in [10], [9] and [3]. These allow the design of OPFB controllers using, e.g., the MATLAB LMI toolbox [7]. However several problems are still open. Most of the solution algorithms are hard to implement and may have restricted solution procedures such as the initial stabilizing gain requirements.

H-Infinity design has played an important role in the study and analysis of control theory since its original formulation in an input-output setting [24]. State-space solutions were rigorously derived for the linear time-invariant case that required solving several associated Riccati equations [4]. Later, more insight into the problem was given after the linear control problem was posed as a zero-sum two-player differential game [2]. An excellent treatment of design is given in [13], which also considers the case of OPFB using dynamic feedback.

A solution to the static output feedback problem is given by Gu and Misra [11] for a restricted class of systems in terms of a single ARE and a related negative definiteness condition. Geromel and Peres [8] give a sufficient condition for a stabilizing static output feedback in terms of a single ARE with a free parameter matrix, plus a condition on the form of the gain matrix. A computational algorithm is proposed. Kucera and de Souza [14] provide necessary and sufficient conditions for stabilizing static output feedback control using

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_24,

these constructions. Gadewadikar et al. [5] extend these conditions to obtain necessary and sufficient conditions for the existence of an H-Infinity static output feedback control. It is shown that the H-Infinity approach can be used for static OPFB design to yield a simplified solution procedure that only requires the solution of two coupled matrix equations. This explains and illuminates the results in [14].That is, design provides more straightforward design equations than optimal control, which requires solving three coupled equations. It is shown that static output feedback H-Infinity solution does not generally yield a well defined saddle point for the zero sum differential game.

Today's industrial systems have complex and possibly unknown dynamics, unknown disturbances, and nonlinear effects including friction and deadzone, and multiple dynamics with different time scales. Such systems include VLSI manufacturing positioning system, hard disk drives, high speed robotic assembly systems, CNC machine tools. Implementation experiences on electromechanical system are addressed in many forums, Kuljaca *et al.* [12] proposed a novel neural network backstepping controller for application to an industrial motor drive system with a high real time computational platform implementation. This paper explains how H-Infinity results [5] can be easily implemented on electromechanical systems. The paper is organized as follows- next two sections present output-feedback design techniques including recent H-Infinity results with numerical design algorithms, description about optimal output-feedback controller and the algorithm is presented for completeness. Section IV elaborates implementation platform with mass spring damper test bed. Last, we compare the results of SVFB, optimal OPFB, with H-Infinity OPFB setup presented in [6]. We have three objectives. First, we give necessary and sufficient conditions for OPFB with design. Second, we suggest a less restrictive numerical solution algorithm with no initial stabilizing gain requirement. Third we compare the performance of implemented controllers.

II. Linear Quadratic Regulator With Output-Feedback

This section presents Linear Quadratic Regulator with Output Feedback. Optimal OPFB control is prescribed in terms of three coupled equations [16].

A. System Description and Definitions

Plant equation is given by the linear time-invariant state variable model

$$\dot{x} = Ax + Bu\,,\; y = Cx\,, \tag{1}$$

The control will be the output-feedback of the form given by

$$u = -Ky = -KCx\,, \tag{2}$$

and a performance index that satisfies

$$J = \frac{1}{2}\int_0^\infty (x^T Qx + u^T Ru)dt\,, \tag{3}$$

with $Q^T = Q \ge 0$, and $R^T = R > 0$. It is assumed that C has a full row rank, a standard assumption to avoid redundant measurement. It is well known that the Optimal OPFB gain is determined by solving the three coupled matrix design equations [16]. The optimal gain design equations are as follows.

$$\begin{aligned}&(A - BKC)^T P + P(A - BKC)\\ &+ C^T K^T RKC + Q = 0,\end{aligned} \tag{4}$$

$$(A - BKC)S + S(A - BKC)^T + X = 0 \tag{5}$$

$$K = R^{-1}B^T PSC^T (CSC^T)^{-1} \tag{6}$$

where $X = E\{x(0)x^T(0)\}$, and S is a symmetric $n \times n$ matrix of lagrange multipliers. An algorithm for solving these is given by [17] and [16]. This algorithm needs an initial stabilizing OPFB gain which is usually difficult to find.

B. Optimal Output-Feedback Algorithm

The equations for P, S, and K are coupled matrix equations in three unknowns. A numerical algorithm is used to find the iterative gains

1. Initialize:

Set $n = 0$, choose initial output-feedback stabilizing gain K_0 such that $(A - BK_0C)$ has stable poles.

2. n-*th* iteration

Solve for S_n

$$\begin{aligned}&(A - BK_nC)S_n +\\ &S_n(A - BK_nC)^T + X = 0\,.\end{aligned} \tag{7}$$

Solve for P_n

$$\begin{aligned}&(A - BK_nC)^T P_i + P_i(A - BK_nC)\\ &+ Q + C^T K_n{}^T RK_nC = 0\end{aligned} \tag{8}$$

Evaluate

$$\Delta K_n = R^{-1}[B^T P_nS_nC^T](CS_nC^T)^{-1} - K_n \tag{9}$$

Set

$$K_{n+1} = K_n + \alpha\Delta K_n \tag{10}$$

α is chosen so that $A - BK_{n+1}C$ is asymptotically stable, and $J_{n+1} \equiv \frac{1}{2}tr(P_{k+1}X) \le J_n$.

If J_{n+1} and J_n are close enough to each other, go to 3 otherwise, set $k = k+1$ and go to 2.

3. Terminate

Set $K = K_{n+1}$, $J = J_{n+1}$.

III. H-Infinity Static Output Feedback Control

In this section we present a method for finding OPFB gains with two equations, no initial stabilizing OPFB gain is needed.

A. System description and definitions

Consider the linear time-invariant system of Fig. 1 with control input *u(t)* output *y(t)*, and disturbance *d(t)* given by

$$\dot{x} = Ax + Bu + Dd\,,\; y = Cx\,, \tag{11}$$

$$x \in R^n\,, u \in R^m\,,\; y \in R^p\,,$$

and a fictitious performance output *z(t)* that satisfies

$$\|z(t)\|^2 = x^T Qx + u^T Ru\,, \tag{12}$$

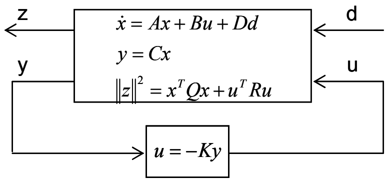


Fig. 1. System Description

By definition the pair *(A, B)* is said to be *stabilizable* if there exists a real matrix K such that $A - BK$ is (asymptotically) stable. The pair *(A, C)* is said to be *detectable* if there exists a real matrix L such that $A - LC$ is stable. System (11) is said to be *output feedback stabilizable* if there exists a real matrix *K* such that $A - BKC$ is stable.

B. Bounded L_2 Gain Design Problem

The System L_2 gain is said to be *bounded or attenuated by* γ if

$$\frac{\int_0^\infty \|z(t)\|^2 dt}{\int_0^\infty \|d(t)\|^2 dt} = \frac{\int_0^\infty (x^T Qx + u^T Ru)dt}{\int_0^\infty (d^T d)dt} \le \gamma^2 \tag{13}$$

Constant output-feedback control is given by Eq 2. It is desired to find a constant OPFB gain *K* such that the system is stable and the L_2 gain is bounded by a prescribed value γ.

Theorem 1. Necessary and Sufficient Conditions for H-infinity Static OPFB Control:

Assume that (A , $\sqrt{Q}$) is detectable. System (11) is output-feedback stabilizable with L_2 gain bounded by γ , if and only if

$$KC = R^{-1}(B^T P + L), \tag{14}$$

where $P = P^T \ge 0$, is a solution to

$$\begin{aligned} &PA + A^T P + Q + \\ &\frac{1}{\gamma^2} PDD^T P - PBR^{-1}B^T P \\ &+ L^T R^{-1} L = 0 \end{aligned} \tag{11}$$

C. H-Infinity Output-Feedback Algorithm

Most existing iterative algorithms for OPFB design require the determination of an initial stabilizing gain, which can be very difficult for practical electromechanical systems. The following algorithm is proposed to solve the two coupled design equations in Theorem 1. Note that *it does not require an initial stabilizing gain* since, in contrast to Kleinman's Algorithm, and the algorithm of Moerder and Calise, it uses a Riccati equation solution, not a Lyapunov equation, at each step.

1. Initialize:

Set *n=0,* $L_0 = 0$, and select γ , Q and *R.*

2. *n-th* iteration:

solve for P_n in

$$\begin{aligned} &P_n A + A^T P_n + Q + \frac{1}{\gamma^2} P_n DD^T P_n \\ &- P_n BR^{-1}B^T P_n + L_n{}^T R^{-1} L_n = 0 \end{aligned} \tag{12}$$

update L

$$L_{n+1} = (B^T P_n + L)C^+ C - B^T P_n \tag{13}$$

If L_{n+1} and L_n are close enough to each other, go to 3 otherwise set $n = n+1$ and go to 2.

3. Terminate:

Set

$$K = R^{-1}(B^T P_n + L_n)C^+ \tag{14}$$

Note that this algorithm uses well-developed techniques for solving Riccati equations available, for instance, in MATLAB

IV. Implementation

We implemented control algorithms on an actual spring mass damper system. Platform simplifies implantation of algorithms and allows a high level of real time system monitoring and interaction with the user.

A. Plant Dynamic Model

The schematic of the system is shown in Fig. 2

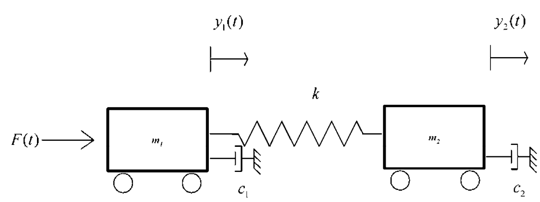


Fig. 2. Two degree Of Freedom Spring Mass System

$$\begin{aligned} &m_1\ddot{y}_1(t) + c_1\dot{y}_1(t) + k(y_1(t) - y_2(t)) = F(t) \\ &m_2\ddot{y}_2(t) + c_2\dot{y}_2(t) + k(y_2(t) - y_1(t)) = 0 \end{aligned} \tag{15}$$

The states and output are assigned as $x = [y_1 \quad \dot{y}_1 \quad y_2 \quad \dot{y}_2]^T$, $y = [y_1 \quad y_2]^T$ respectively. The state-space description is

$$\dot{x} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{k}{m_1} & c_1 & \frac{k}{m_1} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{k}{m_2} & 0 & \frac{-k}{m_2} & \frac{-c_2}{m_2} \end{bmatrix} x + \begin{bmatrix} 0 \\ \frac{k_{hw}}{m_1} \\ 0 \\ 0 \end{bmatrix} u, \tag{16}$$

$$y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} x$$

The control algorithms were implemented on mass-spring damper system described by (15). The parameter values for the case dealt with: $m_1 = 2.77$ kg, $m_2 = 2.59$ kg, $c_1 = 2.1$ N/(m/s), $c_2 = 1.2$ N/(m/s), $k = 830$ N/m, $k_{hw} = 12478.20102$ N/m, hardware gain translates force units into the servo voltage actuation signal. The electrical drive motor dynamics was neglected in the controller design since it is much faster than actual dynamics of the mechanical system. The algorithms were implemented using a real time implementation platform described in the next two sub-sections.

B. Implementation Platform

The user interface to the system is via a PC window environment which supports a broad range of controller specifications, trajectory generation, data acquisition, and plotting features. The experimental control system setup and real time implementation platform is shown in Figs. 3 and 4 respectively.

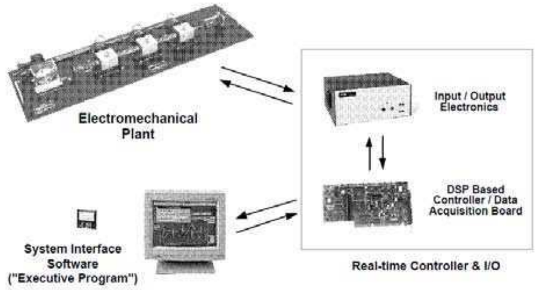


Fig. 3. The Experimental Control System

The spring mass damper apparatus represents many such physical plants including rigid bodies, flexibility in linear drives, gearing and belts. The experimental subsystem is comprised of three subsystems namely electromechanical plant, Real time controller with I/O, and System interface software. Real-time controller unit contains the digital signal processor (DSP) based real time controller, servo/actuator interface, servo amplifier and auxiliary power supplies. The DSP is capable of executing control laws at high sampling rates allowing the implementation to be modeled as being continuous in time. General form is the equivalent numerical form of the given algorithm realizable by the DSP controller board. DSP board executes the algorithm at the specific sample rate. This involves reading the reference input and feedback sensors (optical encoders) values, computing the algorithm, and outputting the digital control effort signal to the digital-to-analog converter (DAC).The DAC converts the resulting stream of digital words to an analog voltage which is transformed to a current by the servo amplifier and then to the torque to a force. The mechanism transforms the motor input to motion at the desired output according to the plant dynamics. The plant outputs are sensed by the encoders which output a stream of pulses. The pulses are decoded by a counter on the DSP board and made available as a digital position word to the real time control algorithm.

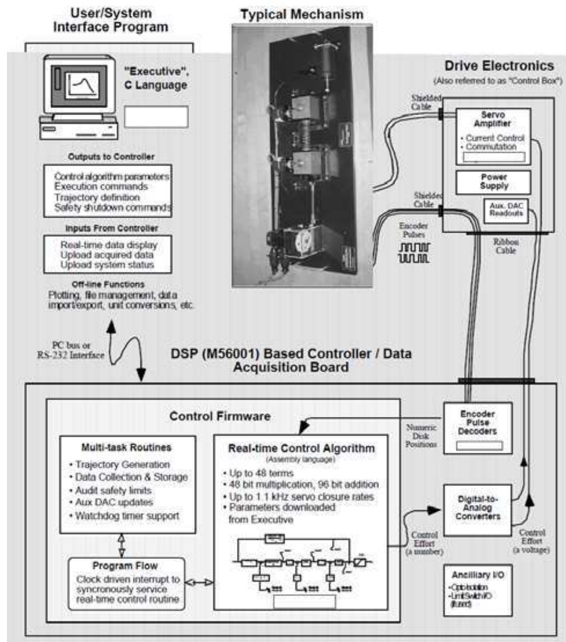


Fig. 4. Real Time Implementation Platform

Full state feedback linear quadratic regulator is implemented, the two degree of freedom configuration with four states is used, and the states are chosen to be mass positions and the rates as per the model described by the state space equations.

V. IMPLEMENTATION RESULTS

Our objective is to do the performance analysis with the disturbance present in the system. Friction is always present in servo-mechanisms and is responsible for tracking lags, steady-state errors, and undesired stick-slip motion. In general, inaccuracies due to friction cannot always be compensated by high gain or by the use of integral action. A damper attached to mass 2 is used to introduce the disturbance in the system.

System is stabilized using three methods for comparison: bounded L_2 gain OPFB, optimal State feedback and Optimal OPFB. Algorithm 1 was used to calculate the bounded L_2 gain OPFB and standard algorithms [16] were used to compute the optimal SVFB and OPFB gains. It is emphasized that algorithm 1 requires no initial stabilizing OPFB gain while determining the optimal OPFB gain was difficult due to the need for finding an initial stabilizing OPFB gain. To see the effectiveness of the design algorithms, system performance is tested for unity step reference input [15]. Fig. 5 describes the closed-loop structure.

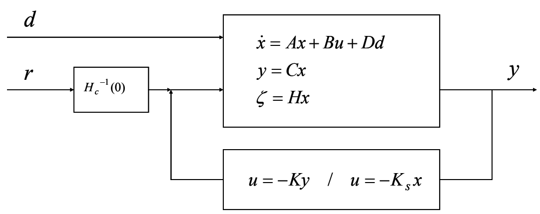


Fig. 5. Closed loop Controller Structure

Fig. 6 summarizes the static OPFB and state feedback controller structures. The second terms in control signals are feed-forward terms added to achieve correct steady-state value of the cart 2 position. Tracking output variable ζ

$$\zeta = Hx = y_2 \quad (17)$$

dictates the formulation of the feed-forward term.

	STATE FEEDBACK	OUTPUT-FEEDBACK
Controller Gains	$K_s = [k_{11} \quad k_{12} \quad k_{13} \quad k_{14}]$	$K = [f_{11} \quad f_{12}]$
Closed-loop transfer function	$H_{cs}(s) = H\{sI - (A - BK)\}^{-1} B$	$H_c(s) = H\{sI - (A - BKC)\}^{-1} B$
Control signal	$u = -K_s x + H_{cs}^{-1}(0) r$	$u = -Ky + H_c^{-1}(0) r$
Closed loop description	$\dot{x} = (A - BK)x + BH_{cs}^{-1}(0) r + Dd$	$\dot{x} = (A - BKC)x + BH_c^{-1}(0) r + Dd$

Fig. 6. State and Output-feedback Controllers

For the computation of bounded L_2 gain output feedback K it is necessary to select Q and R. H-Infinity static output feedback controller gains are obtained with $\gamma = 10$ .Using the algorithms described in earlier section for the given γ , Q and R, the control signal K can be found easily using MATLAB in a few seconds. If the time response and closed loop poles are not satisfactory, the elements of Q and R can be changed and design repeated. After repeating the design several times we selected the design parameters as

$$Q = C_s^{\,T} C_s = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.004 & 0.02 & 2\times10^{-5} \\ 0 & .02 & 1 & 0.001 \\ 0 & 2\times10^{-5} & 0.001 & 1\times10^{-6} \end{bmatrix},$$

and $R = 100000$.

Where $C_s = [0 \quad .02 \quad 1 \quad .001]$ defines the output taken primarily as mass 2 position with a 2 and .1% weighing in the rates of the mass 1 and mass 2 respectively. Note that the tracking output variable state x_3 is weighted heavier than other elements in matrix Q. This Q was selected based on a performance output $y_s = C_s x$. This design shows that the feedback output $y = Cx$ and the performance output can be different. See [15].

For optimal OPFB gain calculation, initial stabilizing gain was taken as $K_0 = [1 \quad 0.1]$. We have assumed that initial states are uniformly distributed on unit sphere, such that $X = I$ [15].

The resulting time simulation responses shown in Fig. 7 compare bounded L_2 gain OPFB, optimal OPFB, and optimal state feedback control. In the simulations all three controllers have commensurate settling times, State-feedback has better transient response characteristics than both bounded L_2 gain output feedback and optimal Output feedback. However, bounded L_2 gain output feedback is superior for practical applications since the actual system has unmodelled dynamics and disturbances that are better controlled by H_∞ Control than by optimal design techniques. Infact, Fig. 8 shows the implementation results. This figure clearly shows that in the actual system L_2 gain OPFB outperforms both the other methods.

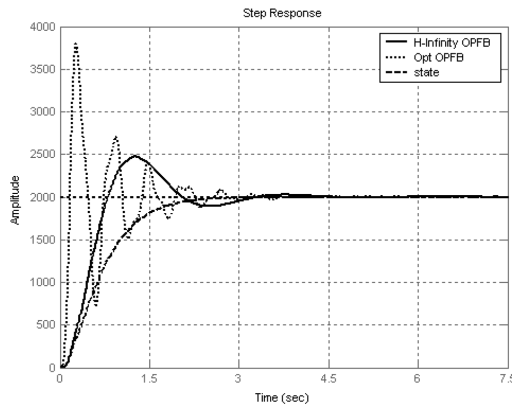


Fig. 7. Simulation Results

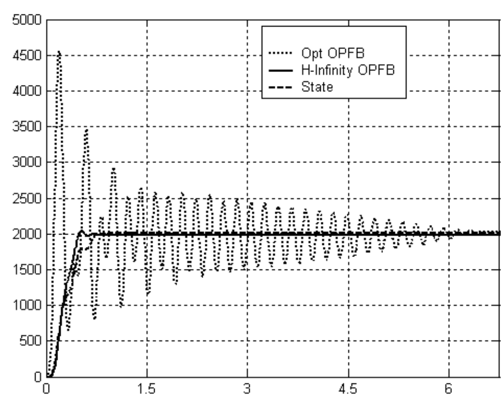


Fig 8. Implementation Results

VI. Conclusion

The Paper presents a complete procedure for design and implementation of structured controllers on an electromechanical system test bed. In this paper performance analysis of L2 gain OPFB, Optimal OPFB, and State-feedback Controllers on a mass-spring system test bed is done. Necessary and sufficient conditions for H-Infinity OPFB are derived. A computational algorithm for the L2 gain OPFB is presented. Implementation results show that H-Infinity OPFB gives superior results when disturbance is present.

REFERENCES

[1] A. Astolfi and P. Colaneri (2000). Static output feedback stabilization of linear and nonlinear systems. Proc. IEEE Conf. Decision and Control, Sydney, 2920-2925.

[2] T. P. Basar , and P. Bernard (1991) H∞ Optimal Control and Related Minimax Design Problems, Berlin, Germany, Birkhauser.

[3] Y. Cao, Y. Sun, and W. Mao (1998). A new necessary and sufficient condition for static output feedback stabilizability and comments on " Stabilization via static Output Feedback". IEEE Transactions on Automatic Control, 43(2), 1110-1112.

[4] J. H. Doyle, K. Glover, P. Khargonekar, and B. Francis (1989). State solutions to standard H2 and H∞ Control Problems. IEEE Transactions on Automatic Control, 34(8), 831-847.

[5] J. Gadewadikar, F. L. Lewis, and M. Abu-Khalaf (2006). Necessary and Sufficient Conditions for H∞ Static Output-Feedback Control. Journal of Guidance Dynamics & Control, AIAA, Vol. 29, No. 4, July–August 2006, pp. 915-920.

[6] J. Gadewadikar, Ajay Bhilegaonkar, F. L. Lewis (2007), Bounded L2 Gain Static Output Feedback: Controller Design and Implementation on an Electromechanical System, Industrial Electronics, IEEE Transactions on, Oct. 2007, Volume: 54, Issue: 5, 2593-2599.

[7] P. Gahinet, A. Nemirovski, A. Laub, & M. Chilali LMI Control Tool box (Natick, MA, Mathworks, Inc, 1995).

[8] J. C. Geromel, and P.L.D. Peres (1985). Decentralised load frequency control. IEE Proceedings, 132(5), 225-230.

[9] J. C. Geromel, C.C. de Souza and R. E. Skelton (1998). Static output feedback controllers: stability and convexity. IEEE Transactions on Automatic Control, 43(1), 120-125.

[10] L. E Ghaoui., F. Oustry, & M. AitRami, A cone complementarity linearization algorithm for static output-feedback and related problems, IEEE Trans. Automat. Contr., vol. 42, 1997, 1171-1176.

[11] G. Gu and P. Misra (1994). J. Guidance, Control, and Dynamics, 17(1), 145-152.

[12] O. Kuljaca, N. Swamy, F. L. Lewis, C. M. Kwan "Design and Implementation of Industrial Neural Network Controller Using Backstepping"

[13] H. W. Knobloch, A. Isidori, and D. Flockerzi (1993). Topics in Control Theory, Birkhauser, Berlin, Germany.

[14] V. Kucera, and C.E. de Souza (1995). A necessary and sufficient condition for output feedback stabilizability. Automatica, 31(9), 1357-1359.

[15] F. L. Lewis (1992). Applied Optimal Control, Digital Design and Implementation. Prentice Hall.

[16] F. L. Lewis, V. L. Syrmos (1995). Optimal Control, 2nd edition, Wiley-Interscience.New York.

[17] D. D Moerder, and A. J. Calise (1985). Convergence of a Numerical Algorithm for calculating optimal output feedback gains. IEEE Transactions on Automatic Control, 30(9), 900-903.

[18] T. R. Parks, "Manual for Model 210/210a Rectilinear Control Systems" Educational Control Products.

[19] B. L. Stevens, and F. L. Lewis (2003) Aircraft Control and Simulation, 2nd ed, Wiley Interscience, New York.

[20] Syrmos, V. L., Abdallah, C.; Dorato, P., "Static output feedback: a survey," Proceedings of 33rd IEEE Conference on Decision and Control, Orlando, FL, 1994, pp. 837-842.

[21] A. Trofino-Neto, and V. Kucera (1993). Stabilization via static output feedback. IEEE Transactions on Automatic Control, 38(5), 764-765.

[22] K. Yang, R. Orsi, and J.B. Moore (2004) A projective algorithm for static output feedback stabilization. Symposium on System, Structure and Control,Oaxaca, (pp 263-268).

[23] J.-T. Yu (2004). A convergent algorithm for Computing Stabilizing Static Output-Feedback Gains. IEEE Transactions on Automatic Control, 49(12), 2271-2275.

[24] G. Zames, Feedback and Optimal Sensitivity: Model reference Transformations, Multiplicative Seminorms, and Approximate Inverses, IEEE Trans. Automat. Contr., vol. 26(2),1981,301-320.

Design of Nonlinear Frequency Controller for Isolated Thermal Power System with Generation Rate Constraint

Ognjen Kuljaca, Jyotirmay Gadewadikar, Kwabena Agyepong
Systems Research Institute, Department of Advanced Technologies, Alcorn State University
1000 ASU Drive #360, Alcorn State, MS 39096, USA
okuljaca@alcorn.edu, jyo@alcorn.edu, kwabena@alcorn.edu

Key-Words: -power system, neural network, adaptive control, frequency control

Abstract: - A nonlinear control scheme for thermal power system frequency control is described. Nonlinear controller is developed through describing function stability analysis of the system Controller ensures stability of the system for any input. Simulation results are given showing the smooth operation of described control scheme. Paper describes in detail the stability analysis and controller design.

1 Introduction

The paper deals with the frequency controller for isolated thermo power system with generation rate constraint. Frequency control becomes more and more important as power systems enter the era of deregulation. It is becoming very hard, if not impossible to schedule loads precisely, thus the load fluctuations in the power system are becoming more explicit. Emerging markets of ancillary services mean that primary controllers and turbines that are used in secondary control change constantly. These changes can cause serious problems when conventional control schemes are used, including problems with stability, unless the primary turbines and controllers are carefully selected. Many ancillary generators sport generation rate constraint as their feature and if secondary controller parameters are not carefully designed presence of generation rate constraint can easily cause the system to become unstable. Here we suggest the controller design procedure for systems with generation rate constraint that would avoid instability under any realistic operating conditions. Design is performed using describing function approach.

The literature about frequency and load – frequency control is numerous ([1], [2], [3], [4], [5], [6], [7], and many others). Many non-adaptive schemes are given ([1], [2], [3], [4], [5] and [6]). Neural Networks (NN) load-frequency control is described in [8], [9] and [10]. The results obtained by using NN controllers are good. However, the described controllers require training. The neural network control scheme that does not require training phase is given in [11]. However, none of these control schemes deals with the systems with generation rate constraints. There are serious problems with using neural networks to control systems with saturation due to requirement for smooth continues functions that describe the controlled process in most cases. One way to deal with this problem is to use non-adaptive fuzzy control ([12], [13]). Non-adaptive fuzzy control showed better performance than conventional PI controllers. However, design of the controller is heuristic and time consuming.

The paper is organized as follows. In Section 2 is given a short preview of describing function approach. A model of isolated thermo power system with generation rate constraint is given in Section 3. System stability analysis using describing function is given in Section 4. Sections 5details the controller design and in Section 6 are given the simulation results. Finally, in Section 7 is given the conclusion.

2 Short Preview of Describing Function Analysis

Describing function is an equivalent gain of nonlinear element, defined by the harmonic linearization method of nonlinear static characteristic ([14], [15]). We can decouple linear and nonlinear parts of the nonlinear system (Fig. 1). If the linear part of the system is a low-pass filter, and if we apply sinusoidal signal to the system,

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_25,

output signal will be of the same frequency as input signal, with damped higher frequencies.

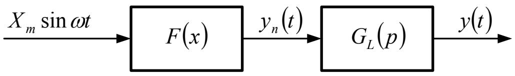


Fig. 1: Nonlinear system represented with decoupled nonlinear $F(x)$ and linear $G_L(p)$ parts, $p = d/dt$

If the amplitudes of higher harmonics are small, relatively to the amplitude of the first harmonic, output signal can be approximated with the first harmonic of the signal. Fourier series expressions for the first harmonic of the signal are:

$$y(t) = Y_{P1} \sin \omega t + Y_{Q1} \cos \omega t \tag{1}$$

$$Y_{P1} = \frac{1}{\pi} \int_0^{2\pi} F(X_m \sin \omega t) \sin \omega t \, d(\omega t) \tag{2}$$

$$Y_{Q1} = \frac{1}{\pi} \int_0^{2\pi} F(X_m \sin \omega t) \cos \omega t \, d(\omega t) \tag{3}$$

Describing function can be expressed as a complex gain element:

$$G_N(X_m) = P(X_m) + jQ(X_m) = |G_N(X_m)| e^{j\varphi_N} \tag{4}$$

where:

$$P(X_m) = \frac{Y_{P1}}{X_m}; \quad Q(X_m) = \frac{Y_{Q1}}{X_m}. \tag{5}$$

Expressions of the Fourier series coefficients can be used for determination of describing function only for simple nonlinear elements.

Describing function can also be determined by the experimental method ([16], [17]). Experimental method includes simulation of the nonlinear element for the range of sinusoidal input signals with changing amplitude and frequency. This method can be used for any nonlinear element, but it may be time consuming.

3 Isolated Thermopower System with Generation Rate Constraint

The model of an isolated thermo power system is shown in Fig. 2.

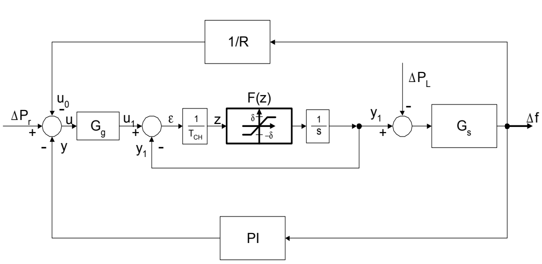


Fig. 2: The model of isolated thermo power system with generation rate constraint

The transfer functions in the model are:

$$G_g = \frac{1}{1 + s \cdot T_g}, \tag{6}$$

$$G_s = \frac{K_s}{1 + s \cdot T_s}, \tag{7}$$

where G_{g},$_t$ T_{CH} and G_s are representing turbine governors, control turbines time constant and the power system respectively. Such models are described in more details in [1], [2], [3], [4], [5], [6], [7], and many others. The generation rate constraint is represented with saturation type nonlinearity shown in Fig. 3

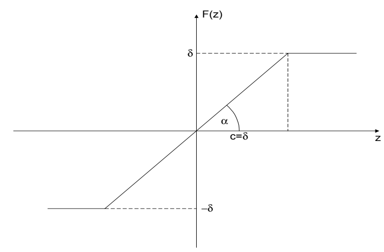


Fig. 3: Generation rate constraint (α = 1)

4 System Stability Analysis Using Describing Function

In order to be able to use describing function in stability analysis, we have to check first if there is a possibility of oscillations in autonomous system (i.e. inputs into system are zero). Dynamics of autonomous systems shown in Fig. 2 with respect to nonlinear part of the system z(t), and without PI controller, can be described with the following nonlinear differential equation:

$$\begin{aligned} &\left[T_G \cdot T_S \cdot p^3 + (T_G + T_S) \cdot p^2 + p\right] \cdot z(p) + \\ &T_{CH} \cdot F(z) \cdot \left[T_G \cdot T_S \cdot p^2 + (T_G + T_S) \cdot p + 1\right] + \\ &\frac{T_{CH}}{R} \cdot K_S \cdot F(z) = 0 \end{aligned} \tag{8}$$

The possible periodic solution of equation (8) has the form $z(t) = Z_m \sin(\omega t)$. Because for our system filter hypothesis stands (linear part of the system is actually low pass filter), equation (8) can be linearized by harmonic linearization procedure ([14], [15] and [18]):

$$\begin{aligned}&\left[T_G \cdot T_S \cdot p^3 + (T_G + T_S)\cdot p^2 + p\right]\cdot z(p) +\\&T_{CH}\cdot G\cdot z(p)\cdot\left[T_G\cdot T_S\cdot p^2 + (T_G+T_S)\cdot p\right] +\\&T_{CH}\cdot G\cdot z(p)\cdot\left[1+\frac{1}{R}\cdot K_S\right] = 0\end{aligned} \tag{9}$$

where G = $G(Z_m)$ is describing function of nonlinearity F(z) (Fig. 3).
From (9) we can find the characteristic equation:

$$\begin{aligned}&T_G\cdot T_S\cdot p^3 + (T_G\cdot T_S\cdot T_{CH}\cdot G + T_G + T_S)\cdot p^2 +\\&\left[T_{CH}\cdot G\cdot(T_G+T_S)+1\right]\cdot p +\\&T_{CH}\cdot G\cdot\left(1+\frac{K_S}{R}\right) = 0\end{aligned} \tag{10}$$

By substitution p = jω in (10) we obtain the complex equation (11).

$$\begin{aligned}&-j\omega^3\cdot T_G\cdot T_S -\\&(T_G\cdot T_S\cdot T_{CH}\cdot G + T_G + T_S)\cdot\omega^2 +\\&\left[T_{CH}\cdot G\cdot(T_G+T_S)+1\right]\cdot j\omega +\\&T_{CH}\cdot G\cdot\left(1+\frac{K_S}{R}\right) = 0\end{aligned} \tag{11}$$

Solutions of (11) that will cause oscillatory mode are:

$$\omega^2 = \frac{T_{CH}\cdot G\cdot(T_G + T_S + 1)}{T_G\cdot T_S} \tag{12}$$

$$G = \frac{1+\frac{K_S}{R}}{T_G + T_S + 1} - (T_G + T_S) \tag{13}$$

For any realistic power system G is a negative value. For nonlinearity F(z) describing function is given as

$$\begin{aligned}&G = G(Z_m) = g(A_m) =\\&\frac{2\cdot tg\alpha}{\Pi}\cdot\left(\arcsin\frac{c}{A_m} + \frac{c}{A_m}\cdot\sqrt{1-\frac{c^2}{{A_m}^2}}\right); A_m > c\end{aligned} \tag{14}$$

which is always positive. Thus, autonomous system cannot establish oscillatory modes. In that case we can check for the stability of autonomous system using any linear system stability criterion. If we use Hurwitz criterion, then the system is stable if

$$\begin{aligned}&(T_G\cdot T_S\cdot T_{CH}\cdot g(A_m)+T_G+T_S)\cdot\left[T_{CH}\cdot g(A_m)\cdot(T_G+T_S)\right] +\\&(T_G\cdot T_S\cdot T_{CH}\cdot g(A_m)+T_G+T_S) >\\&T_G\cdot T_S\cdot T_{CH}\cdot g(A_m)\cdot\left(1+\frac{K_S}{R}\right)\end{aligned} \tag{15}$$

Again, this is always the case for any realistic thermopower system. Thus, autonomous, uncontrolled system (PI controller gains are 0) depicted in Fig. 2 is asymptotically stable for every $g(A_m)$. That means that we can use linear systems stability criterions to establish the stability of the given nonlinear system with PI controller.
Let us now introduce PI controller with transfer function (16),

$$G_r = K_{PR} + \frac{K_I}{s} \tag{16}$$

and let us then define the gain K_o in order to simplify calculations.

$$K_O = K_{PR} + \frac{1}{R} \tag{17}$$

Then, dynamics of the system from Fig. 2 can described, with respect to z(t), as:

$$\begin{aligned}&T_{CH}\cdot z +\\&\left(K_O + \frac{K_I}{s}\right)\cdot\frac{K_S}{1+s\cdot T_S}\cdot\frac{1}{1+s\cdot T_G}\cdot\frac{F(z)}{s} +\\&\frac{F(z)}{s} = 0\end{aligned} \tag{18}$$

By harmonic linearization nonlinearity F(z) is substituted with its describing function:

$$F(z) = G\cdot z; G = G(Z_m) = g(A_m) \tag{19}$$

From (18) and (19) we obtain the characteristic equation of the system:

$$\begin{aligned}&\left(K_O + \frac{K_I}{s}\right)\cdot\frac{K_S}{1+s\cdot T_S}\cdot\frac{1}{1+s\cdot T_G}\cdot\frac{G}{s} +\\&T_{CH} + \frac{G}{s} = 0\end{aligned} \tag{20}$$

We will use Hurwitz stability criterion to determine the stability boundaries of this system. Hurwitz criterion applied on this system states the following requirements on stability of the system (with the respect to K_{pr})

$$m = (T_S \cdot T_{CH}) \cdot (T_S \cdot G + T_G \cdot G + T_{CH}) + (T_G \cdot T_{CH}) \cdot (T_S \cdot G + T_G \cdot G + T_{CH}) + (T_S \cdot T_G \cdot G) \cdot (T_S \cdot G + T_G \cdot G + T_{CH}) - T_S \cdot T_G \cdot T_{CH} \cdot G \cdot \left(\left(K_{PR} + \frac{1}{R} \right) \cdot K_S + 1 \right) > 0 \quad (21)$$

and

$$\left[\left(K_{PR} + \frac{1}{R} \right) \cdot K_S + 1 \right] \cdot m - K_I \cdot K_S \cdot (T_S \cdot T_{CH} + T_G \cdot T_{CH} + T_S \cdot T_G \cdot G)^2 > 0 \quad (22)$$

From (21) and (22) we obtain the graphic representation of the stability regions (Fig. 4).

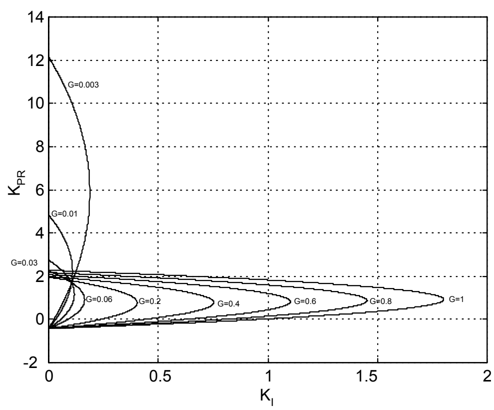


Fig. 4: Stability regions $K_{PR} = f(K_I, G)$

The obvious problem is the choice of the controller gains because they are also functions of G, which is in turn function of reference input and disturbance in the system. In next Section we will show how to design a nonlinear controller using previously established stability regions.

5 Controller Design

The nonlinear controller will apply the stability regions derived in previous Section in order to change proportional gain of PI controller so that system is always stable, notwithstanding the value of G. Value of G can be easily measured and transmitted to secondary control center. If we fix the value of PI integral gain K_I, then we can use (21) and (22) to find an appropriate gain K_{PR} for every value of G. Let us define the following parameters:

$$a = T_S \cdot T_{CH} + T_G \cdot T_{CH} + T_S \cdot T_G \cdot G \quad (23)$$

$$b = T_S \cdot G + T_G \cdot G + T_{CH} \quad (24)$$

$$c = T_S \cdot T_G \cdot T_{CH} \cdot G \quad (25)$$

$$d = K_I \cdot K_S \cdot (T_S \cdot T_{CH} + T_G \cdot T_{CH} + T_S \cdot T_G \cdot G)^2 \quad (26)$$

Then, if we fix the value of integral gain K_I, proportional gain K_{PR} has to satisfy to following conditions for system to be stable:

$$K_{PR} \geq 0 \quad (27)$$

$$K_{PR} < \frac{\frac{ab}{c} - 1}{K_S} - \frac{1}{R} \quad (28)$$

$$K_{PR} < \frac{\frac{Gab + \sqrt{(Gab)^2 - 4Gcd}}{2Gc} - 1}{K_S} - \frac{1}{R} \quad (29)$$

$$K_{PR} > \frac{\frac{Gab - \sqrt{(Gab)^2 - 4Gcd}}{2Gc} - 1}{K_S} - \frac{1}{R} \quad (30)$$

Now, we have designed a controller where at every step K_{PR} is calculated to be between values given in equations (27) to (30). The gain changes depending on the value of G and the system is always stable. Changing gain does not cause problems in the system operation since all the linear parts are low pass filter and there will be no chattering as a result of changing gains.

6 Simulation Results

The simulations were performed for the following set of parameters: T_G = 0.08 s, T_{CH} = 0.3 s, K_s = 120 pu/s, R = 2.4 Hz/pu, Ts = 20 s, δ = 0.017 and K_I =0 .02. The frequency response for step disturbance ΔP_L = 0.01 pu is shown in Fig. 5.

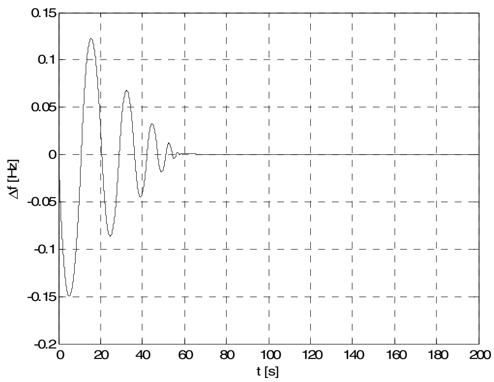


Fig. 5: Frequency change for step disturbance $\Delta P_L = 0.01$

Controller leads the system towards zero frequency error after disturbance. The response is smooth and continuous co changing proportional gain did not affect to expected behavior of the system.

Turbine governor control signal u_1 is shown in Fig. 6. It can be seen that signal is smooth. Only place where are fast changes of the signal are present is after the output of secondary controller. In modern power systems secondary controllers as well as turbine governors are controlled PLC or microprocessors, thus this signal will not have an effect on actual hardware.

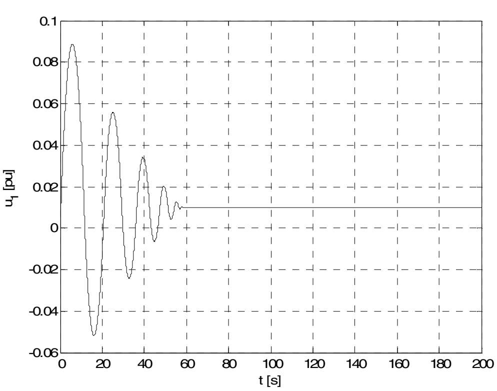


Fig. 6: Turbine governor control signal for step disturbance ΔP_L = 0.01

7 Conclusion

The results of an initial research in designing simple nonlinear controller for thermal power system with generation rate constraint are shown. The controller guaranties stability in any realistic working condition. Nonlinear controller is developed from conventional proportional integral controller by setting proportional gain to change depending on value of amplitude of the signal before nonlinear part of the system. The proportional gain changes in such way that the stability of the system is ensured. Equations required to calculate proportional gain a are simple algebraic terms and can be easily realized with every modern hardware. Simulation results show that system response is smooth and asymptomatically stable..

References:

[1] Elgerd, O.I. "Control of Electric Power Systems", IEEE Proceedings, pp. 4-15, 1981

[2] Elgerd, O.I., Fosha, C.E., "Optimum Megawatt-Frequency Control of Multiarea electric Energy Systems" IEEE Transactions Power Apparatus and Systems, PAS-80, No. 4, pp. 556-562, 1970

[3] Fellachi, A., "Optimum Decentralized Load Frequency Control" IEEE Transactions on Power Systems, Vol. PWRS-2, No. 2, pp. 379-384, 1987

[4] Fosha, C.E., Elgerd, O.I., "The Megawatt-Frequency Control Problem: A New Approach Via Optimal control Theory" IEEE Transactions Power Apparatus and Systems, PAS-80, No. 4, pp. 563-577, 1970

[5] Hamza, M.H., Agathoklis, P., Chan, W.C., "A Combined Selftuning and Integral controller for Load Frequency Control of Interconnected Power systems "Regelungstechnik 30. Jahrgang Heft 7 pp. 226-232, 1982

[6] Kundur, P., Power System Stability and Control. McGraw-Hill, Inc.,1993

[7] Liaw, C.M., "Design of a Reduced-Order Adaptive Load-Frequency Controller for an Interconnected Hydrothermal Power System "International Journal of Control, Vol. 60, No. 6, pp. 1051-1063, 1994

[8] Demiroren, A., Zeyneigly, H.I., Sengor, N.S., "The application of ANN technique to load-frequency control for three-area power system", 2001 IEEE Porto Power Tech Conference, Porto, Portugal

[9] Bevrani, H. "A novel approach for power system load frequency controller design", Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES , Volume: 1 , 6-10 Oct. 2002

[10] Birch, A.P., Sapeluk, A.T., Ozveren, C.T., "An enhanced neural network load frequency control technique", International Conference on Control, 1994. Control '94. Volume 1, 21-24 Mar 1994

[11] Kuljaca, O., Lewis, F. Tesnjak, S., "Neural Network Frequency Control for Thermal Power

Systems", IEEE Control and Decision Conference, 2004
[12] Kuljaca, O., Tesnjak, S., Kuljaca Lj., "Influence of Fuzzy Logic Based Controller on Stability of Isolated Termo Power System Considering Generation Rate Constraint" Proceedings of the 3rd International Conference New Trends in Automation of Energetic Processes, pp. 239-243, Zlín 1998
[13] Sijak, T., Kuljaca, O., Kuljaca, Lj., Tesnjak, S., " Design of Fuzzy Regulator for Power System Secondary Load Frequency Control ", Proceedings of 10th Mediterranean Conference on Control and Automation - MED2002, published on CD (no page numbers), Lisbon 2002
[14] B.N.Novogranov, Determination of Frequency Domain Characteristics of Nonlinear Systems, Masinostroenie, Moskow 1986, (in russian)
[15] Z.Vukic, Lj.Kuljaca, D.Donlagic, S.Tesnjak, Nonlinear Control Systems, Marcel Dekker, New York, Basel 2003
[16] O.Kuljača, S.Tešnjak and Z.Vukić, "Describing Function of Mamdani Type Fuzzy Controller with Input Signals Derived From Single System Input and Singleton Ouput Membership Functions", Proceedings of the 1999 IEEE Hong Kong Symposium on Robotics and Control, Volume I, HKSRC '99, Hong Kong, 1999, pp. 327-331.
[17] T.Šijak, O.Kuljača, and S.Tešnjak, "Stability analysis of fuzzy control system using describing function method", *MED'01*, Dubrovnik Croatia, 26-29.06.2001
[18] Suevalov, L. F., Spravochnik po raschetam sudovyh avtomaticheskih sistem. Sudostroenie, Leningrad, 1977

Designing a Virtual Reality (VR) Storytelling System for Educational Purposes

Siew Yung Lau, Chwen Jen Chen
Faculty of Cognitive Sciences and Human Development
Universiti Malaysia Sarawak
94300 Kota Samarahan Malaysia

Abstract – **This paper addresses the instructional design of a virtual reality (VR)-based educational storytelling system. Indeed, the importance of storytelling in education has long been recognized. Stories are pervasively used as powerful tools in teaching and learning. Although there are efforts to utilize advanced technology, such as virtual reality (VR), in storytelling, most of the research studies focus on immersive-VR that involve expensive and bulky settings, which eventually limit its feasibility to be ubiquitously used. Thus, this project looks into the potential of non-immersive VR, also known as desktop VR, for educational purposes. This alternative low-cost and affordable VR technology requires only conventional computer settings but is able to present an interactive real-time three-dimensional virtual environment (VE) that the learners can navigate in and interact with. A constructivist instructional design theoretical framework [18] is adopted to construct the VR storytelling system. This storytelling system aims to deliver fire safety knowledge to the targeted learners.**

I. Introduction

Stories are pervasively used as a powerful and promising educational means for teaching and learning. Stories draw learners' attention and thus can convey certain messages more easily to them. There are some popular story-based techniques widely used for educational purposes, such as case studies, critical incidents, role playing, and simulations [1]. Through stories, we create and discover meaning in a more engaging manner than just by telling statements of fact. Storytelling is effective as an educational tool because storytelling eases the recall of facts [2], facilitates comprehension, aids in critical thinking [3] and draws attention and concentration [4].

The power of stories in teaching and learning is exhibited from narrative perspectives. Human understanding does not occur as concepts, but as stories organized in narrative form [5]. Storytelling emulates the real life process, where explicit and procedural knowledge are constructed from our daily life. Thus, storytelling is potential in facilitating constructivist learning [6], where learners construct knowledge actively in a learning environment via stories.

Basically, there are two types of storytelling, namely linear and non-linear storytelling (interactive storytelling) [7]. The main distinctive characteristic of these two types of storytelling is the process of the storytelling itself. This project will focus on the non-linear storytelling, where concepts and ideas are the main stem of the storytelling process, without knowing the definite development of the story. Non-linear storytelling is more similar to real life whereby the narrations do not always proceed episodically. Non-linear storytelling allows the audience to have the right to decide how the story goes.

There are researchers who have employed storytelling method in teaching and learning. Kelleher, Pausch and Kiesler [8] show the effectiveness of storytelling in supporting the learning of computer programming. Casey et al. in [9] also show the effectiveness of storytelling via block-building activities in providing an effective context for teaching spatial content. Palmer, Harshbarger and Koch [10] claim storytelling as a constructivist model for developing language and literacy as storytelling is a rich interactive process that facilitates imagination, creative thinking, language abilities, and cooperative learning.

Realizing the potential of storytelling in facilitating teaching and learning, this paper aims to address the instructional design of an educational storytelling system that employs the three-dimensional VR technology. This paper also reviews the current approaches of storytelling, and the potential of non-immersive VR, which is also known as desktop-VR, in facilitating learning via storytelling.

II. Current Approches of Storytelling

A. Non-Virtual Reality (VR)-based Storytelling

The platforms of storytelling are very broad, ranging from oral storytelling, books, music, theatres, movies to technology-enhanced storytelling systems. Apart from conventional storytelling approaches, current storytelling methods are being enhanced with technologies, for instance, electronic multimedia-based storytelling systems [6], which are gaining wide acceptance. There are also various novel approaches of storytelling. In [11], the storytelling process is carried out via multimedia physical storytelling technology, where the participants are placed in a pre-setup room with all the instruments, including the physical objects (which appear in a particular story), sensors, and the physical icons as actuators. However, this type of storytelling requires bulky settings. It also requires the participants to physically explore in the pre-designed limited region and interact by using gross motor (mostly hands). Although it does increase the attractiveness, engagement and the educational benefits of storytelling, it may not be feasible to be used pervasively.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_26, © Springer Science+Business Media B.V. 2010

B. Virtual Reality (VR)-based Storytelling

Virtual reality is used for various purposes, such as entertainment, education, etc. Although there are efforts to use VR technology in storytelling, most of them focus on the use of immersive-VR [12] or mixed-reality system [13]. Besides, the recent research focus is mostly on the construction of attractive and exciting content of the interactive stories [14], intuitive ways of interactivity within the VE as well as sensory feedback and real-time manipulation of the non-linear storytelling [15]. Research also focuses on technical issues such as immersion, interaction, navigation, etc.

Immersive VR requires expensive and bulky settings, which eventually limits its feasibility to be ubiquitously used. Hence, the paper looks into the potential of non-immersive VR, also known as desktop VR, for educational storytelling purposes. Non-immersive VR system presents the interactive real-time three-dimensional VE on conventional computer screens, and the user can interact with it by using generic input devices, such as mouse or keyboard. The fact that no additional peripherals are introduced has made such VR setting to be relatively low cost and thus, affordable to be used by most educational settings or even for personal use.

III. The Educational Potential Of Desktop VR Storytelling System

Desktop-VR is recognized as a technology breakthrough that is powerful in facilitating learning process [16]. Desktop-VR also possesses distinctive capabilities which well support storytelling. These capabilities include:

(1) Visualization of three-dimensional representation of a story episode through various scenarios, which emulates real life environments. This reduces the user's cognitive load if compared to two dimensional representations such as storybooks and multimedia or digital based storytelling systems.

(2) Free navigation in the VEs allows flexible flow of story. As mentioned earlier, the narrations of non-linear storytelling do not always progress episodically and it allows the users to decide the storyline. The fact that VE allows free navigation has indeed facilitated this non-linear storytelling.

(3) Active interface in which the virtual objects or virtual characters behave in accordance to the user's action or state of simulation. This capability supports a dynamic storytelling as the objects or characters in a story are always active.

(4) Interactivity within the VEs also facilitates non-linear storytelling in which the user can decide the storyline flexibly.

Realizing these educational potential, this paper focuses on the instructional design considerations to create an effective VR-based educational storytelling system.

IV. Instructional Design of A VR-based Educational Storytelling System

Instructional design is a systematic and reflective process to plan or design methods of instruction based on principles of learning in order to facilitate the construction of intended knowledge by the learner [17]. Instructional design is important to ensure the learning processes achieve its optimization. Thus, a good instructional design makes learning more effective and efficient.

The pedagogical perspectives that form the theoretical basis of VR-based educational storytelling system include the enhancement of constructivism learning via storytelling, and interactive and explorative learning environment supported by VR technology. This VR-based educational storytelling system employs the instructional design theoretical framework (Fig. 1) proposed by Chen, Toh and Wan in [18]. This model proposes instructional methods that are based on various theories and learning principles to ensure the educational goals are best attained in a VR learning environment.

In general, this model incorporates the concept of integrative goals [19] with the model for designing constructivist learning environments [20], which serve as the macro-strategy of the model. Macro-strategy concerns with the selection, sequence and organization of presenting subject. The framework starts from the inner ring (objectives) and progresses to the outmost ring (support tools) gradually. Generally, the objectives (innermost ring) comprise of a list of individual learning objectives, which intended to be achieved. Then, an integrative goal is derived according to the relationships among the objectives.

This is followed by enterprising scenarios/ problems based on three integrated components: the problem context, the problem representation and the problem manipulation space. The scenarios or episodes of the stories are constructed in this stage. The aggregation of different scenarios forms the non-linear storytelling. The storyline experienced by each user will be different based on their navigation and interaction within the VE. This is followed by designing the necessary tools to support constructivist learning in the VE. These support tools include related cases, information resources and cognitive tools.

This model is enhanced with multimedia design principles derived from the cognitive theory of multimedia learning [21], which serves as micro-strategy of the model. Micro-strategy concerns the strategies to present the learning content effectively. This framework employs constructivist paradigm, which is found to be compatible with the non-linear storytelling. Thus, this framework is used to guide the design of the educational storytelling learning environment.

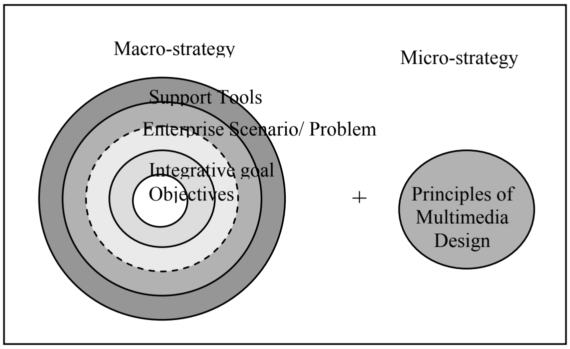


Fig. 1. Instructional design theoretical framework for desktop VR-based learning environment (Adapted: Chen, Toh & Wan, 2004)

V. VR-based Educational Storytelling System on Home Fire Safety

The framework discussed in the previous section is further illustrated with reference to an on-going project on VR-based educational storytelling system on fire safety. According to the National Fire Prevention Association [22], about 396,000 home structure fires were reported in 2006, resulting in 2,580 deaths and 12,500 injuries in the United States. Direct damage had caused an estimated total of $6.8 billion in the same year. A surprising statistic shows that home fires have caused 80% of civilian deaths and 76% injuries. This is a huge proportion from overall death cause. Thus, it is crucial to provide adequate education to the nation about fire safety as prevention is always better than reaction. This section provides an elaboration on how the instructional design theoretical framework is used to guide the design of the storytelling system for this particular learning context, focusing on fire safety at different compartments of a house such as the living room, kitchen, bedroom, and study room.

To begin, the individual learning objectives (starting from the innermost ring of the instructional design framework) which are intended to be achieved are listed as follows:

(1) To identify objects that can easily cause fire, such as heat produce objects, flammable objects, explosive objects, etc;
(2) To identify situations that may cause fire;
(3) To name the various types of fire extinguisher ;
(4) To learn the procedure for using a fire extinguisher;
(5) To recognize different types of fire and the best way to react in time;
(6) To identify self-protection and the ways to escape during fire;

With the objectives stated above, an integrative goal (the second ring) which relates all the individual objectives is identified: The learners should know how to prevent home fire and react to it if occur. Then, scenarios or problems which enable the learners to acquire the knowledge are constructed (the third ring). The flexible story content or the non-linear storyline will serve as the problem context. The problem representation is in speech form (uttered by the virtual character), text form, graphic form and narrative form. The virtual house itself will serve as the problem manipulation space, which includes four main locations: living room, kitchen, bedroom and study room. Support tools (the outmost ring) aid the knowledge construction process; include the activation of the objects or characters behavior which acts as the related cases, the link to extra information such as facts about fire acts as the information resources, and the provided hints which guided the users through the story acts as the cognitive tools. Another instance of cognitive tools is a map of the virtual house provided which enable the learner to be aware of their current location in the virtual house and ease their navigation to the other locations.

In this storytelling system, coaching is used as an instructional activity to give feedback to the learner's performance or actions as the story develops. There are various types of coaching such as providing motivational prompts (e.g. give credit to the learners for performing a good task), monitoring and regulating the learners' performance (e.g. asking the learner to check out the ambiguous "zhi...zhi..." sound when they approach the multi-socket) and provoking reflection (e.g. asking the learners whether they place the candle properly). Procedural scaffolding is another instructional activity which is used to scaffold the learners' learning process in the VE. The "Help" button will link to the page that provides information on navigation in the VE.

Apart from the macro-strategy discussed previously, micro-strategy also plays its significant role in enhancing the learning which is guided by principles of multimedia. In this storytelling system, relevant images, sounds and labels are incorporated to enhance the interpretation and learning process, for instance, a fire image in attached to the button "Facts of Fire". Fig. 2 shows a screenshot of the educational storytelling system.

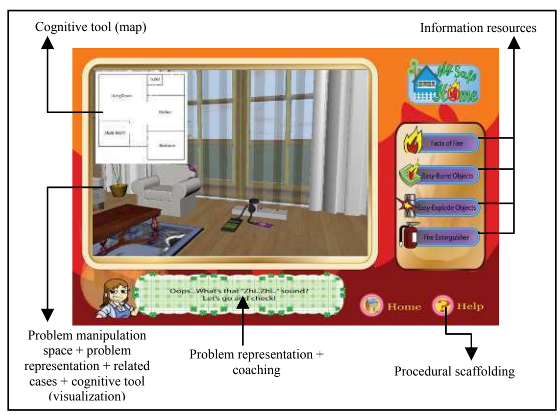


Fig. 2. Screenshots of the educational storytelling system

In the VE, learners are guided through stories where they are able to navigate to a particular location and the VE will be activated and some scenarios are created which require them to perform some tasks and at the same time convey the message of fire safety to them. However, the learners still have the right to navigate to anywhere else in the virtual house. Hence, unlike conventional storytelling method, this VR-based storytelling system allows the learners to have a flexible flow of story while constructively create knowledge on home fire safety.

VI. CONCLUSION

This paper presents the potential of using storytelling approach in a desktop-VR based learning environment. It also suggests the feasible use of a theoretical framework that can be utilized by instructional designer or educator to guide the instructional design of such learning environment in the effort to produce effective and engaging educational storytelling process.

REFERENCES

[1] K. Taylor, C. Marienau, and M. Fiddler, *Developing Adult Learners.* San Francisco: Jossey-Bass, 2000.

[2] Y. George, and B. Schaer, "An investigation of imposed-induced imagery methods on kindergarten children's recall of prose content", in *Proceedings of Annual Meeting of the Mid-South Education Research Association*, vol. 48, 1986.

[3] N.K. Aiex, "Storytelling: its wide-ranging impact in the classroom", *Eric Digests*, No. 9, 1988.

[4] A. Baker, and E. Greene, *Storytelling: Art and Technique*. New York: Bowker, 1997.

[5] J. Bruner, Acts *of Meaning*. London: Harvard University Press, 1990.

[6] B.C. Palmer, S.J. Harshbarger, and C.A. Koch, "Storytelling as a constructivist model for developing language and literacy", *Journal of Poetry Therapy*, 14(4), 2001, pp. 199-212.

[7] K. Björk, "Online Storytelling", *E-content Reports by ACTeN*, May, 2004.

[8] C. Kelleher, R. Pausch, and S. Kiesler, "Storytelling Alice motivates middle school girls to learn computer programming", *CHI 2007*, San Jose, California, USA, April 28–May 3, 2007.

[9] B.M. Casey, N. Andrews, H. Schindler, J.E. Kersh, A. Samper, and J. Copley, "The development of spatial skills through interventions involving block building activities", *Cognition and Instruction*, 26(3), July 2008, pp. 269-309.

[10] M. Mazyrah, W.A. Wan Fatimah, M.N. Shahrina and S. Suziah, "A conceptual framework for English language courseware using storytelling approach: Case study in University Teknologi Petronas", in *Proceedings of International Symposium on Information Technology*, August 2008.

[11] J. Montemayor, A. Druin, G. Chipman, A. Farber and M.L. Guha, "Tools for children to create physical interactive StoryRooms", *ACM Computers in Entertainment*, vol. 2, No. 1, January 2004, Article 3.

[12] M. Cavazza, J. Lugrin, and D. Pizza, " Madame Bovary on the Holodeck: Immersive Interactive Storytelling", *MM'07*, Ausburg, Bavaria, Germany, September, 2007.

[13] J. Cassell, M. Ananny, A. Basu, T. Bickmore, P. Chong, D. Mellis, K. Ryokai, J. Smith, H. Vilhjalmsson and H. Yan, "Shaared Reality: Physical Collaboration with a Virtual Peer", *ACM CHI 2000, Conference Proceedings*, The Hague, The Netherlands, 2000.

[14] L. Schäfer, "Storytelling in Virtual Environment", *Virtual Reality,* 7(1), 2003.

[15] M. Cavazza and Donikian S. (Eds.), "Virtual Storytelling: Using virtual reality technologies for storytelling", 4th International Conference, ICVS 2007, Saint-Malo, France, December 5-7, 2007.

[16] D. Fällman, A. Backman, and K. Holmlund, "VR in Education: An introduction to Multi-sensory constructivist learning environment", *Conference on University Pedagogy*, Umea University, Sweden, 1999.

[17] P.L. Smith and T.J. Ragan, *Instructional Design*. John Wiley & Sons, Inc., 1999.

[18] C.J. Chen, S.C. Toh, and M.F. Wan, "The theoretical framework for designing Desktop Reality-based learning environments", *Journal of interactive Learning Resarch*, 15(2), 2004, pp. 147-167.

[19] D.H. Jonassen, "Designing constructivist learning environments", in C.M. Reigeluth (Ed.), *Instructional-design Theories and Models: A New Paradigm of Instructional Theory*, vol. 2, New Jersey: Lawrence Erlbaum Associates, 1999, pp. 215-239.

[20] R.M. Gagne and M.D. Merrill, "Integrative goals for instructional design", *Educational Technology Research and Development*, 38(1), 1990, pp. 23-30.

[21] R.E. Mayer, *Multimedia learning*. Cambridge University Press, 2002.

[22] *Fast Facts about Fire.* National Fire Prevention Association. Retrieved 15 August 2008 at http://www.nfpa.org

Automated Marking of Transfer Function Problems

David Moore
Department of Electrical Engineering, University of Cape Town, Rondebosch, South Africa

***Abstract* – The increasing significance of e-learning, on-demand lecturing and auto-tutoring systems in tertiary education necessitates automated evaluation systems. Current literature presents a number of applied methods for evaluation, but little specific research on the evaluation of transfer function problems found in control engineering curricula. This paper presents a number of methods for use in the assessment of such mathematical entities, with particular emphasis on its application to problems in system identification. Further, the paper raises a number of concerns that need to be addressed when designing a complete evaluation procedure for dynamic systems.**

***Keyword* – Engineering Education, E-Learning, Educational Software, System Identification, Control Engineering.**

I. Introduction

The current trend of larger classes and limited resources in engineering education necessitates a change in the current program of a typical engineering degree [1] – [3]. The inclusion of software tools is seen as a "*complement to classical teaching and laboratory work*" [2]. Introducing e-learning software to engineering education is a way "*to do more in the same amount of time"* [1]. It allows students to benefit from on-demand teaching, which while present in small classes, is lost in the large class environment [4]. On-demand teaching will require on-demand assessment to reclaim an important characteristic of small class learning environments.

Various systems for on-demand teaching and automated assessment in control engineering [5] – [7] as well as for general mathematics [8] have been described in the literature. These systems address on-demand testing: Providing problems for students to solve, evaluating student answers, and giving rapid feedback to the students.

Previous work on computer based education [4, 9] indicates that automated assessment of large classes is both feasible and useful to students and staff.

In control engineering, [3, 10] stated that modelling and system identification problems are imperative to achieving good system control. Thus, practice and testing of the ability to deduce an adequate dynamic model from a set of plant data is an important skill that needs to be imparted in control engineering education.

The topic of system identification has been addressed through software [4, 6, 7]. These usually result in a solution in the form of a transfer function g(s):

$$g(s) = \frac{b_0 + b\ s + \ldots + b_m s^m}{a_0 + a_1 s + \ldots + a_n s^n} e^{-s\tau} \qquad (1)$$

that is a ratio of polynomials and includes a deadtime term. This could be readily evaluated using a computer algebra system [8], if it weren't for the inexact nature of system identification problems.

In these systems [4, 5], students are expected to apply a number of techniques to noisy, corrupted data in order to identify an adequate system model. This mimics system identification using industrial signals (e.g. Figure 1).

The inherent approximation of the system identification task is the crux of the assessment problem, as there is no exact solution, nor described reasoning behind most of the assessment methods in use [5] – [7].

Thus, the problem of assessing transfer functions in system identification is not the trivial case of comparing an exact solution against a submitted answer. Instead it is measuring how 'close' a student has come to approximating the solution and how to define what 'close' is.

This paper focuses on methods for use in an assessment model for dynamic systems defined by transfer functions (1). It specifically investigates the design of an assessment system for an introductory Control Engineering course.

II. Recent Research

A number of different assessment techniques have been applied to on-demand assessment systems in teaching.

Reference [6] makes use of a "fuzzy" range grading system, where the answer is evaluated against a triangle function with height equal to 100% centred on the solution, and a base width dependent on the accuracy required by the marking. The intention is to give students partial credit for answers close to the correct solution.

Reference [7] looks at frequency domain system identification and design testing. It uses a fairly complex system of assessment, generating multiple different assessment

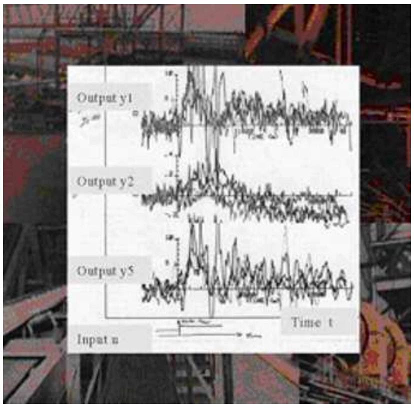


Figure 1 : Typical Plant Signals

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_27, © Springer Science+Business Media B.V. 2010

measures and using a fuzzy logic system to choose the most appropriate one.

Of interest is the different methods used to assess the answers. The first is a cost function measure of the error in the magnitude and phase plots of a Bode plot, which is then marked against a fixed range "*fuzzy*" system similar to that in [6]. The second is a "*measure [of] the relative distance between the singularities from the real and the identified transfer functions*" [7]

Another [8], presents the option of computer algebra systems. These systems are able to directly evaluate fairly complex algebra. While these systems are advancing steadily and are able to assist in the assessment of various algebraic systems including ratios of polynomials, none of those sampled handled the approximate assessment required.

III. Naming Conventions

The following conventions are used:

For the system identification problem, the student's submission is the *answer,* whereas, the *solution* is the exactly correct result.

It is useful to categorize methods by their basic function:

Scaling methods remove ambiguity in the answer by scaling both answer and solution to some value, or part of their transfer functions.

Scoring methods typically convert the answer to a metric representing the total error in the answer or distance from the solution.

Grading methods convert a scoring metric to a marker friendly percentage through a one-to-one mapping.

IV. Assessment Considerations

One of the first issues raised during the investigation of assessment techniques was that of relative versus absolute tolerances in marking. When the answer is only in the vicinity of the actual solution, it is necessary to define the range around the correct answer where marks are assigned, and what accuracy is required.

Relative tolerances scale depending on their order of magnitude, but become undefined around zero.

Absolute tolerances do not suffer from such problems, though they are unsuitable for marking where the order of magnitude of parameters varies greatly. They are however dependent on the subjective nature of marking. This is best illustrated through an example:

In systems identification, the steady state gain, eta, of the dynamic system must be correctly identified, where eta = b_0/a_0.

The correct sign of the gain (positive or negative) is imperative to a correct answer. Yet, around zero an absolute tolerance could evaluate an answer with incorrect sign as correct due to its absolute proximity to the correct answer.

No simple solution was found to fix this, but the best practical approach was to use a combined small absolute and relative tolerance system.

V. Possible Assessment Solutions

There are many ways to assess transfer function answers against known solutions. From the literature it appears that no single overriding method is currently being used. Thus, a number of different possible methods are explored and some of the considerations involved in choosing an evaluation system are presented.

Below is a selection of methods which are considered, split up based on their major characteristics.

Parametric methods approach transfer function assessment by reducing them to a set of defining parameters, which can then be evaluated.

Cost Function methods use integral squared error and similar methods as a means of assessment. These are applied to both time and frequency domain representations.

VI. Parametric Assessment

Parametric methods convert the transfer function (1) into a set of parameters that encapsulate the system.

$$P = \{b_0, b_1, \ldots, b_m, a_0, a_1, \ldots, a_n, \tau\} \tag{2}$$

This parameter vector for an answer can be compared element-by-element with a similarly defined vector for the solution.

As there is no easy way to ensure that students input their answer in exactly the same format as the solution, it is necessary to apply some means of scaling to the representations of the transfer functions before parameterization.

As an example, the two dynamic systems g1 & g2 are identical,

$$g_1(s) = \frac{1}{2+3s} = g_2(s) = \frac{3}{6+9s} \tag{3}$$

yet parameterize to P1 & P2 that are not the same.

$$P_1 = \{1,2,3\} \neq P_2 = \{3,6,9\} \tag{4}$$

This was found to have a significant effect on the resulting mark and thus various methods for scaling are explored.

A. Scaling Methods

Scaling standardizes the representation such that their parameterizations match. This is mostly achieved by multiplying, dividing or removing factors from the representations.

A number of different types of scaling representations were identified for transfer function problems. While not exhaustive, they give a fair representation of the possibilities and their effects.

i. One Variable Scaling

One Variable Scaling uses a single variable to scale the system parameters. The partial monic form is one example that puts the denominator into monic form:

$$g_{pm}(s) = \frac{\frac{b_m}{a_n}s^m + \frac{b_{m-1}}{a_n}s^{m-1} + \ldots + \frac{b_0}{a_n}}{s^n + \frac{a_{n-1}}{a_n}s^{n-1} + \ldots + \frac{a_0}{a_n}} e^{-s\tau} \quad (5)$$

while the partial gain form shows the steady state gain, eta, explicitly in the numerator:

$$g_{pgf}(s) = \frac{\frac{b_0}{a_0} + \frac{b_1}{a_0}s + \ldots + \frac{b_m}{a_0}s^m}{1 + \frac{a_1}{a_0}s + \ldots + \frac{a_n}{a_0}s^n} e^{-s\tau} \quad (6)$$

Although this scaling is simple, it has the disadvantage that it creates a dependence on one of the variable, such that any error in that parameter is propagated onto all the other parameters, and results in a far greater change in the resulting mark.

Figure 2 shows the relationship between error (as a fraction of the value) and the resulting mark. As expected a change in a_0 has a far greater effect than any other variable.

Similar results occur for other One Variable Scalings.

ii. Two Variable Scaling / Factor Scaling

This is similar to One Variable Scaling. It removes a factor from both numerator and denominator. The monic form converts both denominator and numerator to monics:

$$g_m(s) = \frac{b_m}{a_n} \cdot \frac{s^m + \frac{b_{m-1}}{b_m}s^{m-1} + \ldots + \frac{b_0}{b_m}}{s^n + \frac{a_{n-1}}{a_n}s^{n-1} + \ldots + \frac{a_0}{a_n}} e^{-s\tau} \quad (7)$$

The Gain Form extracts the steady state gain as a constant multiplier:

$$g_{gf}(s) = \frac{b_0}{a_0} \cdot \frac{1 + \frac{b_1}{b_0}s + \ldots + \frac{b_m}{b_0}s^m}{1 + \frac{a_1}{a_0}s + \ldots + \frac{a_n}{a_0}s^n} e^{-s\tau} \quad (8)$$

Similar to One Variable Scaling, errors in two of the variables results in a far greater effect on the final mark than the other variables.

iii. Derived Scaling

These scaling methods use subjective or derived values.

Zero-Pole-Gain (ZPG) scaling uses the zeros, poles and gain as the parameterization, as in [7]. The zeroes and poles are represented as complex numbers in the Laplace variable s.

$$g_{ZPG}(s) = \gamma \cdot \frac{(s+z_1)(s+z_2)\ldots(s+z_m)}{(s+p_1)(s+p_2)\ldots(s+p_n)} e^{-s\tau} \quad (9)$$

With parameters:

$$P_{ZPG} = \{\gamma, z_1, z_2, \ldots, z_m, p_1, p_2, \ldots, p_n, \tau\} \quad (10)$$

Eta Time – Constant scaling uses an alternate representation where the constants are the gain and settling times of the poles and zeros.

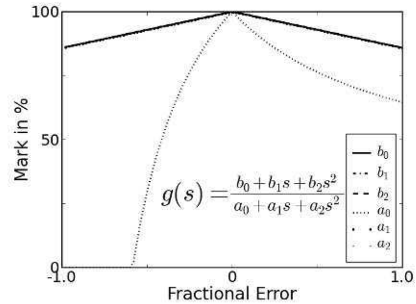


Figure 2: Partial Gainform Scaling

$$g_{ET-C}(s) = \eta \cdot \frac{(1+\beta_1 s)(1+\beta_2 s)\ldots(1+\beta_m s)}{(1+\alpha_1 s)(1+\alpha_2 s)\ldots(1+\alpha_n s)} e^{-s\tau} \quad (11)$$

With parameters:

$$P_{ET-C} = \{\eta, \beta_1, \beta_2, \ldots, \beta_n, \alpha_1, \alpha_2, \ldots, \alpha_n, \tau\} \quad (12)$$

Finally another example of a subjective parameterization is suggested in [4], where the parameters are different values that specify a generic transfer function form.

$$g_b(s) = \frac{\alpha}{\alpha} \frac{b_0 + b_1 s}{s^M (a_0 + a_1 s + a_2 s^2)} e^{-s\tau} \quad (13)$$

With parameters:

$$P = \{m, \alpha b_0, \alpha b_1, \tau, M, n, \alpha a_0, \alpha a_1, \alpha a_2\} \quad (14)$$

where α is defined to optimize the students mark.

B. Scoring Methods

Once the representation has been standardized through scaling, the answer is evaluated against the solution.

Two methods suggested in [4] provide a means to evaluate the parameterizations. Both methods use the normalized Euclidean distance --- The first one for each parameter; the second one for the parameter vector.

The methods provide a summative metric

$$f_s() = \frac{\sqrt{\sum (sol_i - ans_i)^2}}{\sqrt{\sum sol_i^2}} \quad (15)$$

or a formative metric

$$f_f() = \frac{\sum_{i=1}^{N} \sqrt{\frac{(sol_i - ans_i)^2}{sol_i^2}}}{\sum_{i=1}^{N} 1} \quad (16)$$

to measure the error in the answer supplied by the student.

Both of these methods could readily include subjective weighting based on the parameterization used.

Two problems inherent in the formalization of the formative methods were that of zero parameters, and that of significance of parameters.

i. Zero Parameter Solution:

Due to the normalization involved (which effectively adds a relative scaling in both scoring methods), any solution *sol* containing parameters that are zero, will create undefined situations for the formative metric. A reasonable modification to the formula that takes care of this situation, and solves the absolute versus relative tolerance problem presented above is the addition of a unit offset to the denominator of the normalisation, as shown:

$$f_f() = \frac{\sum_{i=1}^{N} \sqrt{\frac{(sol_i - ans_i)^2}{(sol_i^2 + 1)}}}{\sum_{i=1}^{N} 1} \tag{17}$$

ii. Significance of variables

The second issue is that of the denominator sum to N, where N is the length of the largest dimension in either answer or solution vector in the formative method. This effectively gives an even weighting to all parameters in a given parameterization. The problem is that due to scaling and parameterization some of these parameters are either always the same, or not significant.

Thus another reasonable alteration is made on the premise that the marker knows in advance the number of significant variables M. Thus the method becomes:

$$f_f() = \frac{\sum_{i=1}^{N} \sqrt{\frac{(sol_i - ans_i)^2}{(sol_i^2 + 1)}}}{M} \tag{18}$$

For example, it fixes the case in which the student enters a polynomial of larger degree than that contained in the solution.

C. Grading Methods:

As a final step in the assessment, the scoring metric is converted into the more traditional form of a percentage.

Three possible methods for doing this are the *fuzzy range* method [6], the *Gauss* method and the *piecewise fuzzy* method.

These methods take a scoring metric, f (), and a marking tolerance, *mf,* and use these to create a percentage grade *M*.

The *fuzzy* method assigns a mark by a simple triangle function.

$$M = 100 \times \left(1 - \frac{f()}{mf}\right) \geq 0 \tag{19}$$

The *Gauss* method was developed from a Gauss distribution with the intention of providing evaluation information over the entire range of possible answers as opposed to the limited range of the fuzzy marking system. It is intended for use for summative feedback.

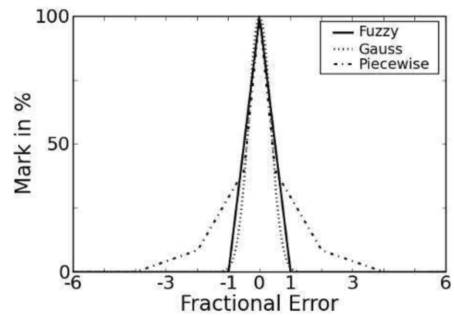


Figure 3: Grading Method Percent vs Error

$$M = 100 \times \exp\left(\frac{-f(\ )^2}{(mf/2)^2}\right) \tag{20}$$

Finally, the piecewise fuzzy method is a balance between the *Fuzzy* and *Gauss* methods. It allows for the wide range of the Gauss system for summative marking, and combines it with the accurate mark determination of the fuzzy system.

$$\begin{aligned} M = 50 \times \left(1 - \frac{f(\ )}{mf/2}\right) \geq 0 + 33 \times \left(1 - \frac{f(\ )}{2 \times mf}\right) \geq 0 \\ \ldots + 17 \times \left(1 - \frac{f(\ )}{4 \times mf}\right) \geq 0 \end{aligned} \tag{21}$$

The selection of each method is greatly dependent on the subjective application of the assessment system and what educators want from it.

VII. Cost Function Methods

Another group of methods identified as possible means for assessment are those that use the time response or frequency domain plots as the representation of the system to be evaluated.

A. Time Domain

Time domain methods use the step or impulse response of the system. This signal can then have various cost functions applied to create an error metric, similar to the scoring step in parametric assessment.

Important factors for time domain simulations are how long to simulate for, and what sampling frequency to use.

B. Frequency Domain

Frequency domain methods investigated used the Bode plot representation of the systems.

Similar costs as those used in the time domain, were used for both magnitude and phase plots. These two measures were then averaged to get a scoring metric.

VIII. System Identification Auto-Tutor:

The above methods were used to design an assessment system for a system identification auto-tutoring program. A number of combined and individual assessment methods were proposed for testing to determine which was the most suitable.

A. *Subjectively Important Measures*

The tutoring program was designed to be an automated exercise for use in a first course in control engineering.

As such the assessment system was to focus on subjectively important concepts. For the purpose of the test program, these were the concepts of steady state gain and pole dominance.

i. *Importance of Steady State Gain*

The steady state (DC) gain of a system is a significant feature that has an overriding effect on the entire system. As such the assessment system needs to ensure that it dominates the effect of other parameters with regards to the resulting mark.

ii. *Pole Dominance*

In control systems the poles and zeros (denominator and numerator roots respectively) of the transfer function, have varying effect (on the dynamics of the represented system) depending on their relative size (or placement).

In the case of the introductory control courses and many industrial applications, the focus is primarily on the effect of the dominant poles and zeros. While less dominant poles and zeros play a role, they do not have as great an impact on the overall system.

For this reason, the assessment system should weight the effect of the dominant poles/zeroes on the resulting mark more heavily.

B. *Method Selection*

Two tests were run to determine the suitability of each method.

i. *Parameter Perturbation*

By parameter perturbation, the first test evaluated each proposed system based on their sensitivity to the two identified subjectively important features. This was done by perturbing individual parameters of a set system g(s) and analyzing the resulting marks.

The assessment systems investigated were:

- A time mean integral absolute normalized error system, shown in Figure 4.

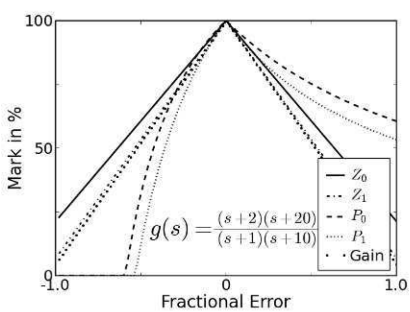


Figure 4: TIANE Error vs Mark in Percemt

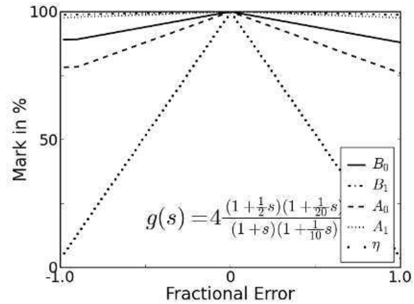


Figure 5: Eta Time-Constant Error vs Mark in Percent

- A frequency magnitude and phase, mean integral absolute normalized error system.
- Both split and combined fuzzy summative eta time-constant scaled systems. Figure 5 shows the desired characteristics of overriding dependence on steady state gain and inherently weights dominant poles heavily.
- Both split and combined fuzzy summative zero-pole-gain systems. Figure 6 shows that the Zero-Pole-Gain parameterization weights the largest parameters most heavily, thus making the dominant pole the least important feature.

The split systems used the summative method separately on the gain (gain or eta), the poles/zeroes or time constants, and the deadtime (tau), and the geometric average for the marks was taken as the final mark.

ii. *Function Correlation*

For the second test, a sample system identification problem was posed to students. The answers collected would be used to test each proposed method. Because of the limited scope of students' submitted answers, additional randomly perturbed answers were added to this set for a more complete range of answers.

The sample problem (22) in *ZPG* form was marked against the student and randomly perturbed answers, and the results analyzed for trends and correlation.

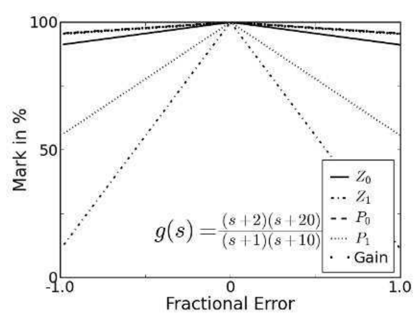


Figure 6: Zero-Pole-Gain Error vs Mark in Percent

$$g(s) = 2 \cdot \frac{(s+2)}{(s+1)} \quad (22)$$

From these tests, it became apparent that certain functions gave relatable marks. Figure 7 and Figure 8 show a correlation between each of the integral methods and a corresponding parametric method.

Figure 7 shows the relation between the Time Cost Function method and the Eta Time – Constant method with a correlation coefficient of 0.96. Similarly Figure 8 shows the relation between the Frequency Cost Function and the Zero-Pole-Gain method with a correlation of 0.6.

IX. Conclusions

From all of the methods reviewed there is no single method that unequivocally is better than the other methods. Certain methods proved more successful at certain types of assessment than others, while some did not perform as intended. There are two different approaches that can be taken to transfer function assessment, the pragmatic approach of practical engineering; where the assessment tries to balance complexity with workable solutions, and the approach of pedagogic assessment, where any deviation from the correct answer should be penalized. This complicates the selection of assessment methods, as each method is suitable to one or other application.

The *parametric* methods proved to be fast and provide a wide range of possible assessment situations.

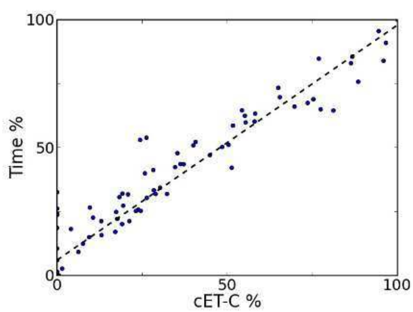


Figure 7: Time Cost Function vs Eta Time-Constant Percentages

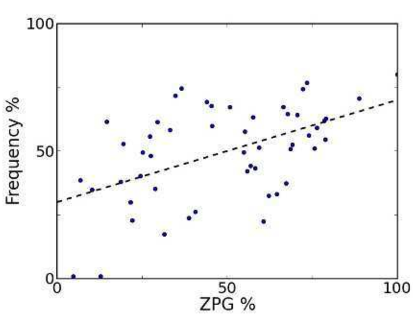


Figure 8: Frequency Cost Function vs Zero-Pole-Gain Percentages

The *Time Domain* cost function methods provided a means for accurate curve fitting assessment, allowing for mismatched models and parameters. The *Frequency Domain* required a strictly correct model form, with the correct number of poles and zeros. Whenever this was met the frequency domain costs gave a good measure of accuracy of the transfer function.

For pragmatic engineering assessment, the *ET-C* and *Time Domain* Cost Function methods provide suitable means for evaluation. Their high correlation indicates that the *ET-C* method can be used over the time domain method due to its relative efficiency and speed.

Similarly for the pedagogic assessment approach, the *ZPG* and *Frequency Domain* cost function methods provide suitable evaluation. The two methods correlate reasonably; yet the *ZPG* method is more efficient than the *Frequency Domain* method.

The various one and two variable scalings proved to be ineffective as the introduced sensitivity to error in specific variables did not consistently relate to their relative effect on the overall system they represented.

For the auto-tutoring program, two assessment models were chosen for implementation. The first, a split *summative fuzzy ET-C* method, would be used for the first year introductory course where pole dominance and steady state gain are key features. The second, a *combined formative fuzzy ZPG* method, would be included for use in later courses in control engineering, where the accuracy of the model is more important than the basic understanding.

Acknowledgments

Thanks to Simon Spicer for his insight and help on all things mathematical.

The project was funded by a grant from CET.

References

[1] W. A. Wulf, "How shall we satisfy the long-term educational needs of engineers?," *Proc. IEEE*, vol. 88, pp. 593–596, 2000.

[2] G. J. C. Copinga, M. H. G. Verhaegen, and M. J. J. M. van de Ven, "Toward a web-based study support environment for teaching automatic control," *IEEE Contr. Syst.*, vol. 20, pp. 8–19, 2000.

[3] S. D. Bencomo, "Control learning: present and future," *Annu. Rev. Control,* vol. 28, pp 115-136, 2004

[4] M. Braae, "An Auto-Tutor for System Identification in Control Engineering," *ICEL*, 2008

[5] R. C. Garcia and B. S. Heck, "Enhancing classical controls education via interactive GUI design," *IEEE Contr. Syst.*, vol. 19, pp. 77–82, 1999.

[6] V. Petridis, S. Kazarlis, and V. G. Kaburlasos, "ACES: An Interactive Software Platform for Self-Instruction and Self-Evaluation in Automatic Control Systems*," IEEE Trans Educ.*, vol. 46, no. 1, pp. 102-110, 2003.

[7] J. F. Haffner, L. F. Alves Pereira, and D. F. Coutinho, "Computer-Assisted Evaluation of Undergraduate Courses in Frequency-Domain Techniques for System Control*," IEEE Trans Educ.*, vol. 49, no. 2, pp. 224-235, 2006.

[8] C. J. Sangwin, "Assessing mathematics automatically using computer algebra and the internet," *Teaching Mathematics and its Applications*, pp. 1-14, 2004, ProQuest Education Journals

[9] M. Braae, N. J. Van Zyl and A. R. Anderson, "Evolutionary Development from simulation teachware towards multimedia and automated education," *IFAC*, 2001

[10] N. A. Kheir, K. J. Åström, D. Auslander, K. C. Cheok, G. F. Franklin, M. Masten, and M. Rabins, "Control Systems Engineering Education," *Automatica*, vol. 32. no. 2, pp. 147-166, 1996.

[11] R. T. Pena, F. G. Jota, and C. Seixas-Filho, "A New Undergraduate Degree in Control Engineering," *IEEE Trans Educ.*, vol. 44, no. 1, pp. 399-405, 2001.

Predicting Persistence of College Freshmen Using Neural Networks

Cameron Ian Cooper
Fort Lewis College
1000 Rim Drive
Durango, CO 81301 USA

Abstract— **Nationally, a significant percentage of public undergraduate students do not successfully go onto their second year of college. Low retention rates incur a heavy cost on both schools and students. Retention rates vary by institution. Some factors that have a role in student retention are school selectivity, geography, academics, athletics, and socio-economics. Finding the students that best fit a given school's overall environment poses an important challenge to schools. Understanding student persistence (i.e., enrolling in the subsequent second semester) offers a first step to understanding retention (i.e., enrolling in the subsequent second year). This research explores the creation of a decision support system that profiles the successful student for a particular school. The system scores individual students based upon their probability of persistence. This paper details a highly accurate (better than 90%) neural network-based decision support system that predicts first-semester persistence. The system offers a means to score an individual applicant for recruitment. The system also offers a detailed profile of the student that is most likely to be a successful student at a particular institution of higher education.**

Index Terms **— Neural Networks, Persistence, Predictive Modeling, Recruitment, Retention**

I. Introduction

As mentioned earlier, a significant percentage, nearly 40%, of college and university freshman do not retain to their second year of higher education [1]. This comes at a time when "the [expected] production of high school graduates will slow moderately" over the next decade [2]. If colleges and universities could provide targeted marketing towards students that are more likely to be a good fit at their particular school, then much of the problem created by low retention rates might be mitigated. Students having a good fit with the school are probably more likely to persist beyond the first semester and therefore more likely to retain beyond the first year. This research offers a framework for a decision support system based upon neural computing that scores individual students/applicants based upon their probability of persisting.

A. Related Research

With the goal of creating a predictive model for first-semester persistence, an effective approach had to be chosen. Many approaches exist to perform predictive modeling (e.g. regression, decision trees, cluster analysis, Bayesian networks, and neural networks). The author chose to build a neural network. Neural computation offers a nontraditional numerical tool for predicting future behavior. Neural networks have been shown to be capable of utilizing complex, nonlinear, and noisy data for predictive purposes [3]. In fact, applications of neural networks have been found to be of use in many diverse fields [4] such as:

- Loan application analysis
- Fraud detection
- Stock advisory
- Routing
- Bond rating

Furthermore, neural networks with at least one hidden layer are universal approximators [5]. A universal approximator is able to a find a functional mapping between inputs and outputs if such a mapping exists. Given the research and the proven utility of neural computation over the past two decades, the author opted to explore neural networks for the prediction of freshman persistence.

Related to persistence and educational outcomes, neural network engineers have shown that neural networks are capable of predicting student success [6,7,8,9]. There is reason to believe a multilayer neural network can be similarly employed to predict first-semester persistence.

II. Methods

A. Data Collection

The site for this research, Fort Lewis College, is a public liberal arts college in Southwestern Colorado. Fort Lewis College is the only four-year college within a 150 mile radius of Durango, Colorado[10]. Similar to other institutions of higher education, Fort Lewis College has a significant portion, 42.1%, of its entering student body, that does not retain to their second year of study [11]. This rate is in line with national norms mentioned earlier, and thus results from this research seen at this school may possibly be generalized to other institutions of higher education.

As mentioned earlier, the utilization of a neural network was chosen as the data mining method for predicting student persistence. Specifically, a multilayer backpropagation neural network was developed using NeuroSolutions as the development platform. A multilayer backpropagation neural network, also known as a multilayer perceptron, is the most

Manuscript received October 15, 2008.
C. I. Cooper is with Fort Lewis College, Durango, CO 81301 USA. (e-mail: cooper_c@fortlewis.edu).

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_28, © Springer Science+Business Media B.V. 2010

widely used neural network paradigm [12]. NeuroSolutions gives neural network engineers a vast and comprehensive array of options in the construction and training (both supervised and unsupervised) of neural networks. In this situation, archived student data can be obtained with known outcomes of persistence for supervised training. Hence, a neural network engineer would most likely opt to create a neural network trained via historic student data.

With a goal to create a database of two prior freshman cohorts with known first-semester persistence, Institutional Research at Fort Lewis College queried relevant input variables from the school's student information database. The Fall 2005 and the Fall 2006 freshman cohorts (n=1768) were, at the time, the most current cohorts with known first-semester persistence. The author chose the relevant variables within the query via interviews with admissions personnel, general education faculty, advisors, the Dean of General Education and Exploratory Studies, and Institutional Research. The resulting initial set of possible predictors had 58 variables which included but was not limited to area such as:

- High school information (e.g., GPA, ranking, class size)
- Standardized test scores
- Geographic location
- Ethnicity
- Financial Aid information
- Gender
- Basic skills requirement upon admission
- Residency
- Tuition rebates

This data became the base training data for the supervised training of the multilayer backpropagation neural network. Institutional Research joined this data with a single column for each student's first-semester persistence in order to make this dataset complete for training.

B. *Input Pruning*

At the end of the data collection phase, the author had a database with 1768 rows of student information across 58 different variables. In order to find the most generalizable decision support solution for predicting persistence in a production setting, this set of variables needed to be pruned down to the most essential and meaningful variables. Thus, before the author engineered and trained a neural network using NeuroSolutions, a simple correlation analysis was performed. Spearman's Rho correlation coefficient was calculated between each of the 58 potential predictors and the binary persistence output variable. Statistically, Spearman's Rho is the most appropriate measure of linear association (positive or negative) given the ordinal level of measure for some of the variables. Using an alpha level of 0.05 (2-tailed)*, 18 of the 58 possible predictors were statistically significant. The author pruned the input space to 20 inputs per industry and the software's recommended "rule of thumb" [13]. All of these inputs had to be known at the time of admission in order for the model to be of use for recruitment purposes. The 20 inputs/variables with their corresponding Spearman's rank correlation coefficient are detailed in the table below:

TABLE I
INPUTS FOUND VIA SPEARMAN'S RHO CORRELATION

Predictor	Description	Rho
Gender	Gender	.092(**)
WUE	From Western Undergraduate Exchange State	-0.035
Out-of-State Native American	Out-of-state Native American Classification	-.146(**)
Out-of-State Resident	Out-of-state Resident	.045(*)
In-state Resident	Colorado Resident	.085(**)
Required remedial hours	Number of calculated remedial hrs at admission	-.121(**)
HS GPA	High school GPA	.167(**)
HS Class Size	High school graduating class size	.075(**)
HS Percentile	High school percentile	.058(**)
Hispanic, Black, and Other	Declared ethnicity	-.048(*)
School of Arts, Humanities, Soc Sciences (AHSS)	Intended major with the school of (AHSS)	.057(**)
School of Business Admin (SOBA)	Intended major in SOBA	-.050(**)
ACT Math	Highest ACT mathematics score	.094(**)
ACT English	Highest ACT English score	.157(**)
ACT Composite	Highest ACT Composite score	.145(**)
CCHE Index	Colorado Commission on Higher Education calculated Index Score [14]	.181(**)
From Local HS	From local high school within 60 mile radius	-0.011
Submitted FAFSA	Submitted a Free Application for Federal Student Aid	.156(**)
Total Need	Total need as determined by FAFSA	-.092(**)
Native American Tuition Waiver	Participant in FLC Nat Amer Tuition Waiver	-.144(**)

Of interest, within the aforementioned 20 predictors, 16 of the predictors are statistically significant at an alpha level of 0.01 (2-tailed)**.

The author included two other predictors, WUE and Local HS, which were shown to not be statistically significant as measured by Spearman's Rho. The inclusion of these variables came at the request of the Director of Admissions at Fort Lewis College. WUE is a binary variable indicating whether or not a student is from a Western Undergraduate Exchange state. These students receive a rebate on their tuition at Fort Lewis College. Local HS is a binary variable indicating whether or not the student came from a local high school (i.e. within a 60 mile radius of the college). The author did not see the problem in including these variables because the next stage of the input pruning process was to consider each variable for elimination via backwards elimination.

Backwards elimination entailed building a neural network with all 20 predictors to determine the base predictive accuracy. Then, starting with the least statistically significant variable, eliminate it and rebuild the neural network and determine if the predictive accuracy increased or not due to the elimination of the variable. WUE and Local HS both ended up improving the predictive accuracy of the neural network. Thus, they were ultimately retained as predictors in the final network. Because a neural network with at least one hidden layer considers nonlinear combinations, the author believes these predictors work in a nonlinear combination with other variable(s) to improve predictive accuracy.

Before building and training the base neural network, however, the data was oversampled to provide more equitable treatment between the persisting and non-persisting categories. At Fort Lewis College, approximately 20% of entering freshman do not persist to the second semester. Within the training data (2005 and 2006 cohorts), 19% of the students did not persist. Once the data was oversampled, the author had 2910 rows of data for training with an equal number of rows for persisting students and non-persisting students. The goal of oversampling was to have approximately equal predictive accuracies from the neural network for both categories.

Finally, in order to score individual students in regards to their probability of persistence, the neural network output was postprocessed using a SoftMaxAxon provided by the NeuroSolutions software package [15]. An initial cutoff point of p=0.50 for categorization was utilized. A student with a probability greater than one half was classified as probable persister, and a student with a probability less than one half was classified as a likely non-persister.

III. RESULTS

Using NeuroSolutions, the author created a multilayer backpropagation neural network with two hidden layers. NeuroSolutions refers to their user interfaces for creating neural networks as "breadboards." The term is borrowed from the field of electrical engineering in reference to the physical prototype of a circuit. The breadboard for the neural network within NeuroSolutions is depicted below. The postprocessing SoftMaxAxon, which transforms the neural network output into a probability can be seen at the bottom right of the breadboard in the figure below.

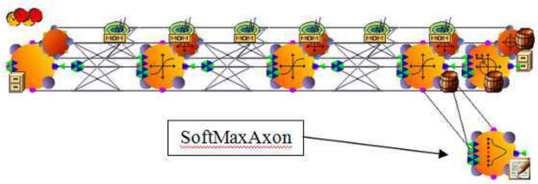


Fig. 1. Breadboard used for network within NeuroSolutions

In training the neural network depicted above, the author chose to utilize the Leave-N-Out training algorithm within NeuroSolutions. By using the Leave-N-Out algorithm, all 2910 rows of data could be used for training. The Leave-N-Out training algorithm, also referred to as dynamic cross validation, allows all of the data to be used for training and testing [16]. The final neural network with two hidden layers appeared to train to an acceptable level of error as indicated by the graph of the mean squared error (MSE) below:

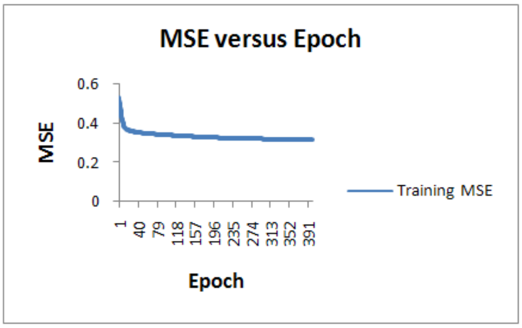


Fig. 2. Mean Squared Error versus Epoch

The MSE of the neural network model shown by above learning curve levels off as the number of epochs increases towards the end of the figure. This suggests that the learning reached a point that further training would be redundant and unnecessary, because it would not improve the overall model. The author believes the final neural network model used for actual prediction was fully trained and suitable for prediction in a production setting.

The final neural network, after the pruning process of backwards elimination, ultimately has 11 predictors with an overall predictive accuracy of 67.9% using the probability cutoff of 0.50 mentioned in the previous section. This compares to the overall predictive accuracy of 64.9% seen before the backwards elimination process. While the time consuming process of backwards elimination only resulted in an overall 3% predictive improvement, the more important finding was a model with only 11 predictors as opposed to 20. Such a model provides a much more generalizable solution that should work well with out-of-sample test data. The final 11 important inputs/predictors are listed in the table below:

TABLE II
INPUTS VIA BACKWARDS ELIMINATION

Predictor
Gender
WUE
In-state Resident
Required remedial hours
HS Class Size
HS Percentile
ACT Composite
CCHE Index
From Local HS
Submitted FAFSA
Total Need

Of interest, within the aforementioned 20 predictors,

As indicated by the icon for the SoftMaxAxon in the breadboard, the SoftMaxAxon attempts to fit the neural network output to a normal distribution in order for the network engineer to be able to interpret the output as a probability. In order to test the final neural network, the most current data with known first-semester persistence (the Fall 2007 cohort) was used as an out-of-sample dataset for testing. The following frequency distribution indicates that the SoftMaxAxon provided output approximately conforming to a normal distribution. Given this output, one can reasonably interpret the final neural network output as a probability of persistence.

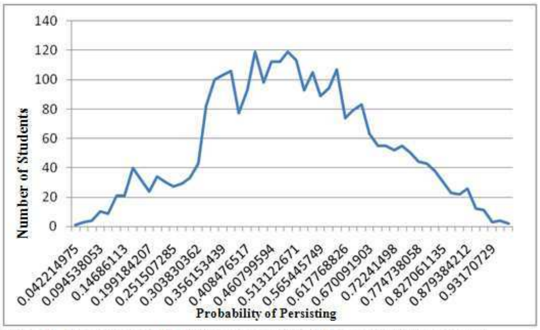


Fig. 3. Frequency distribution output from SoftMaxAxon

Using a cutoff of 0.50, the model identified 351of 925 students as probable persisters. 86.5% of these students actually persisted thus giving an accuracy of 86.5%. Using a cutoff of 0.90 (i.e. at least a 90% probability of persisting), the model identified 134 students as probable persisters. Of these students, 123 persisted for a 91.8% accuracy. The varying cutoffs with their respective accuracy and student count are shown in the graph below.

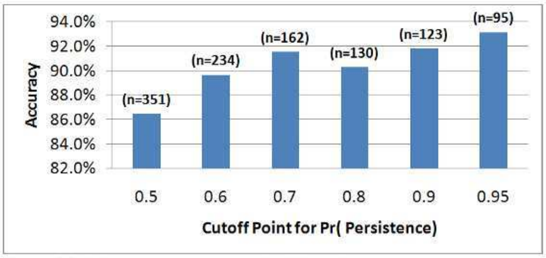


Fig. 4. Model accuracy versus probability of persistence

III. CONCLUSIONS

By using a relatively large (n=925) out-of-sample test dataset (i.e. data not used for training), the author has high confidence in the generalizability of the predictive model detailed in this research. The research offers a highly accurate model for predicting first-semester persistence at a particular institution, Fort Lewis College. This research can be viewed as a framework for other institutions wanting to have a similar predictive model. The author envisions the model to be of great used for admission personnel in the recruitment of students that would be a good fit for a specific institution. Additionally, the identified important predictors collectively offer administrators a profile of a student that is likely to persist. Such a profile can be extremely valuable for targeted marketing campaigns, curricular reform, and bettering overall retention of the school.

Regarding the effective use of this model, additional research needs to be completed. First, a cost benefit analysis needs to be performed in order to determine the most viable and beneficial probability cutoff point. This cutoff will likely vary on an institutional basis. Also unique to the institution, the profile for of a good-fitting student offered by the model can help determine where these students come from and help administrators develop a strategy to get more of them. The scoring capability of the predictive model would again be applied. Additionally, this first-semester persistence model can be viewed as a template for the creation of a first-year retention model. If a school were to have both an accurate first-semester persistence model and an accurate first-year retention model, then school personnel could target the ideal student and then take measures to keep that student beyond the first year and further into his or her higher education. Finally, given a refined model, the first-semester persistence model may be of use in predicting the size of an incoming class for revenue determination purposes.

ACKNOWLEDGMENT

The author wishes to thank Dr. Carol Smith, Dean of Enrollment Services and General Education at FLC for her support of this research and Amy Yeager, director of Institutional Research at FLC for her help in querying data.

REFERENCES

[1] ACT. (2008). National Collegiate Retention and Persistence to Degree Rates. Retrieved October 2, 2008 from http://www.act.org/research/policymakers/pdf/retain_2008.pdf.

[2] WICHE. (2008). Knocking at the College Door: Projections of High School Graduates by State and Race/Ethnicity, 1992 to 2022. Retrieved October 2, 2008 from http://www.wiche.edu/Policy/Knocking/1992-2022/knocking_complete_book.pdf.

[3] Soetanto, R., & Proverbs, D. G. (2004). Intelligent models for predicting levels of client satisfaction. Journal of Construction Research, 5(2), 233-253.

[4] Hagan, M. T., Demuth, H. B., & Beale, M. H. (1996). Neural network design (1st ed.). Boston: PWS Pub.

[5] Haykin, S. S. (1999). Neural networks : a comprehensive foundation (2nd ed.). Upper Saddle River, N.J.: Prentice Hall.

[6] Hardgrave, B. C., & Wilson, R. L. (1994). Predicting graduate student success: a comparison of neural networks and traditional techniques. Computers & Operations Research, 21(3), 249-264.

[7] Naik, B., & Ragothaman, S. (2004). Using neural networks to predict MBA student success. College Student Journal, 38(1), 143-150.

[8] Sexton, R. S., Hignite, M. A., Margavio, T., & Satzinger, J. (2001). Neural networks refined: using a genetic algorithm to identify predictors of IS student success. Journal of Computer Information Systems, 41(3), 42-48.

[9] Cooper, C. I. (2008).A Neural Network-Based Decision Support System for Identifying and Remediating At-Risk Students in Developmental Mathematics.” *International Journal of Applied Management and Decision Sciences, March*.

[10] Fort Lewis College. (2008). Facts at a Glance. Retreived October 6, 2008 from http://www.fortlewis.edu/about_flc/facts_at_glance.asp.

[11] FLC Institutional Research. (2008). Retention and Graduation. Retrieved October 6th, 2008 from http://www.fortlewis.edu/ir/retention.htm.

[12] Turban, E., Aronson, J. E., & Liang, T. P. (2005). Decision support systems and intelligent systems (7th ed.). Upper Saddle River, New Jersey: Prentice Hall.

[13] NeuroDimension. (November 2001). Designing neural networks: selecting inputs. Retrieved October 6, 2008, from http://www.nd.com/newsletter/nov_2001.htm.

[14] CCHE, (2008). CCHE Admissions Eligibility Index. Retrieved October 10th, 2008 from http://highered.colorado.gov/Publications/Policies/Current/i-partf-index.pdf.

[15] NeuroDimension. (May 2007). Interpreting Outputs of a Classification Neural Network as Probabilities. Retrieved October 7, 2008, from http://www.nd.com/newsletter/Tip_Box_v7_2.html.

[16] NeuroDimension. (December 2005). Leave-N-Out Training. Retrieved October 6, 2008, from http://www.nd.com/newsletter/Newsletter_v6_4.html.

Euro-QLIO a New Educational Tool for Specialist Training at European Level

H. Pop, C. G. Alionte, M. Prisecaru, V. Apostol, G. Popescu, I. D. Filipoiu, B. Heit, T. Gueorguiev

***Abstract*—In the context of constant development of methods for distance learning, or e-learning, the paper presents a project in which the software tools specific to this type of education are being applied, namely, the ERASMUS project called „Filière Euro Qualité Logistique des Organisations" , acronym EURO-QLIO. The e-learning process offers flexibility and affordability "anywhere, anytime", so that it can be used in both hybrid and joint learning. The project EURO-QLIO ensures the preparation and conduct of the partnership for training in the field of Metrology, Quality Control and Organization Logistics. The partnership is ensured by the involvement of three renowned universities from three different countries: „University Henri Poincaré" from Nancy, France (which is the project coordinator) "University Politehnica of Bucharest", Romania, and "University Angel Kanchev" from Rousse, Bulgaria. The Romanian partner has co-opted from the industrial area the firm "SC Marco & Alex Instalați Frig SRL". This partner represents the interface between the Euro-QLIO project and the industrial beneficiaries. The educational platform is based on two very important components, namely: establishment of a virtual campus and issuance of a diploma with triple recognition.**

***Index Terms*—distance learning, EURO-QLIO project, joint education, triple recognition**

I. Introduction

Some of the most used software tools are the e-learning platforms. An e-learning platform leads to an effective teaching and learning process, created by involving digital content into a community defined by learning groups or virtual classes, offering support throughout the learning process. Known as online distance learning, based on internet resources, the e-learning offers flexibility and affordability "anywhere and anytime", so that it can be used in both hybrid and joint learning.

The result of the learning process can be customized according to the organization needs and required competences, giving the learner full control and flexibility.

Also the e-learning platforms offer solutions to some important issues which appear in high level education. In this context we can include the need for long term professional education, reduction of investment and time stress over resources, the growth of student diversity and innovation regarding ways of attending the courses: distance "part-time" learning, opened and flexible, and not last the possibility to interconnect different tools used in human resources departments. Using e-learning involves a better learning process through the following: a variety of documentation resources: printed materials, presentations, questionnaires, multimedia files – graphic files, animations, sounds, interactive elements, etc.; offers control over the way in which students learn, being able to study when they want and where they want; the students have an integrated learning environment which can be customized according to personal needs; establishment of an environment that promotes an active way of learning; communication improvement between teachers and students or between students through "tele-presence"; offers continuous or discontinuous reactions; increases student motivation by means of interactive courses and study groups; supports and encourages collaborative teaching; offers economic support by re-using expensive resources such as software applications; gives students full responsibility regarding the learning process; reduction of administrative activity by providing the didactic material online; improves overall teaching process; improved efficiency; cuts out differences between reach and poor people, boys and girls, rural and urban population; cuts out the problems caused by distance; improves time for providing information – lower time spent for transforming new information into didactic material; possibility of synchronic access (at the same time) or at different periods of time to the e-learning platform.

The implementation of e-learning platforms in Romania faces the following difficulties: infrastructure – telecommunication services, access to the Internet or intranets and informational databases; teachers have low computer driving skills; the need to create new jobs in order to support telecommunication, digital information transfer and multimedia applications; publisher rights over the didactic material and the costs of software applications.

II. The Erasmus Project EURO - QLIO

In spring 2007, the papers containing the proposal of the project „Filière Euro Qualité Logistique des Organisations" was submitted in the competition of European Program ERASMUS. This project was built to ensure the preparation and conduct of the partnership for employment training in the field of Metrology, Quality Control and Organization Logistics. The partnership was established through the involvement of three renowned universities from three

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_29, © Springer Science+Business Media B.V. 2010

different countries: „Universitatea Henri Poincaré" from Nancy, France (which is the project coordinator) "Politehnica University of Bucharest", Romania, and" Angel Kanchev University" from Russe, Bulgaria.

The three partner universities provide well prepared "finished product" (human resource), and in order to establish a connection with the industrial beneficiaries, the Romanian partner has co-opted from the industrial area the firm "SC Marco & Alex Instalații Frig SRL". This partner represents the interface between the educational project and the industrial beneficiaries, potential financial resource, ensuring in the same time the future partnership with other commercial societies, employers of human resource provided by the this program.

The project was approved and was launched in October 2007, being included in the ERASMUS Program under the code 134395-LLP-1-2007-1-FR-ERASMUS-EVC, acronym EURO –QLIO. The total budget of the project is 446.733 €, of which 300,000 € funding from the European Community (67,15%) and has 24 months.

The project „Filière Euro Qualité Logistique des Organisations", EURO-QLIO, aims to achieve a platform for distance learning in the field of Metrology, Quality Control and Organization Logistics. This platform is based on two very important components, namely: establishment of a virtual campus, on the one hand, and issuance of a diploma with triple recognition, on the other hand.

The distance learning platform will enhance the students exchanges between the partners, will promote the cooperation between teachers and, thus, will increase the mobility of each partner. Also, an extension of the project by co-opting new partners is intended, so that the EURO-QLIO project becomes a reference at European and international level in terms of education quality.

Virtual campus relies on the use of means from the Information and Electronic Communications Technology (IECT), providing a set of software tools specific to distance learning – integrated multimedia modules used for a joint education, involving both individual and group collaboration. Also, joint education integrates methods of classic teaching and interactive teaching having the possibility of regrouping students in a flexible manner. In this context, the software tools have to be implemented in the teaching system of each partner. The virtual campus is focused on license training, ensuring at the same time the base premises for a complete learning cycle involving both Master and PhD studies in the field of Metrology, Quality Control and Organization Logistics.

For each group of students a tutor from each of the partner universities will be appointed. The tutor will observe the evolution of the students regarding their education and psycho-education.

The diploma with triple recognition is given to the students after they graduate the learning modules coordinated from distance by the teaching teams belonging to each of the partner countries, France, Romania and Bulgaria. The diplomas obtained after going through the different learning stages are recognized by each partner due to the fact that the particularities of the national teaching system are eliminated by introducing virtual mobility both, for students and teachers.

Internationalization of diplomas will be made primarily by: introducing a unitary curriculum in four languages (Romanian, French, English and Bulgarian); organizing the field work in specialized industrial units with a similar profile (industrial logistics) or in the partner universities; software transfer for different applications.

EURO-QLIO project is based on the Information and Electronic Communications Technology for education and professional development. In this regard, the project aims for the following: supporting a permanent link with technology - use of the web platform as the basis of educational process; the material dimension of technology - access to information resources necessary to support the process of education and the social dimension of technology – issues regarding the depersonalization of student-teacher contact trough a joint education approach; a more dynamic process of education - a course does not belong to a person, but to a team of teachers who are improving and modifying it continuously from the curricular and contents point of view.

Considering an analogy between a company and a university, when using the information technology in the education process, is concerned the application of knowledge trough tools, methods, means, rules, standards etc. The result of knowledge application relies on transforming the raw human resource, the high school graduate, into the finished product, the graduate of a high level education. In the industrial area, different ways of processing are becoming more and more automated. Furthermore, this comes in the support of the student because by applying a joint education (traditional and distance education), one person may find much easier to assimilate knowledge and to gain abilities, as he can choose his own way of learning.

By applying the traditional way of teaching, the student can have a dialog with the teacher and colleagues in his native language, during the course and after reading the materials available on the online platform. In case that the support course needs to be modified, it can be done very quickly in each of the four languages. Also, the student can be helped by his colleagues through team work using the software tools.

The information technology used in the EURO-QLIO project, must cover three main functional issues: material issue, represented by machines, tools, instruments, installations etc.; normative issue, represented by rules, design strategies and management; social issues, represented by abilities, capabilities, individual and social behavior.

The teaching activity is designed by a team of teachers formed from all three universities involved in the project, based on the following activities: establishing the teaching objectives; investigation of existing educational resources; setting the team of teachers; establishing the content; applying the methodology; ensuring evaluation of the teaching activity.

Computer assisted learning is a method which uses the principles of traditional teaching in the context of new information and communication technologies. This method is a peculiarity of our modern society, for which the information processing systems offer new methods, features and means to accomplish the teaching process by: computerization of teaching-learning-evaluation activity; improvement of educational process through actions regarding management, documentation, communication and interrogation; automated and interactive simulation of knowledge and abilities engaged in the teaching process according to official documents of educational planning; observing the professional development of students and providing activities of constant improvement; completion and correction of the educational process.

At the same time, the use of IECT methods improves the teaching process through heuristic and individual coordination of teaching-learning-evaluation activities: organizing and re-organizing the teaching process in accordance with the curriculum, adaptable to the abilities of each student; cognitive challenge of the student through joint teaching and evaluation with the role to detect shortcomings and eventual problems; solving the teaching tasks presented above by re-activating or obtaining the required information from information resources reached via computers; achievement of materials and resources in multimedia format used for running through home works, lessons, groups of lessons, chapters or disciplines; providing additional activities to stimulate the student creativity.

Designing the training activity involves management, ordering and correlation of the support course of each of the partner countries. The support course will be taught, learned and evaluated at the functional level between teacher and student, in order to provide the same knowledge and abilities in each of the three universities.

A problem that must be considered, due to heterogeneous teaching processes belonging to each country, is how the teaching process can be achieved: defining the teaching objectives; establishing the content; applying the methodology; ensuring the evaluation of teaching and learning activities. To solve this problem the teachers must communicate between them using the resources provided by the virtual campus and mobility programs, absolutely necessary for leading the work groups.

Designing the teaching process is based on the concept of work package. In fact it is a logic and semantic group of disciplines. Each work package is coordinated, controlled and organized by a responsible and each team involved in a work package aims for research, theorization and implementation of training strategies.

The design activity involves precise methods for development, implementation, evaluation and operation of functional structures which facilitates the learning of scientific subject whatever the complexity of the discipline may be. Through the EURO-QLIO project, these concepts are implemented and verified in order to sketch the teaching system with triple recognition. This teaching system is a combination of instruments meant to help the learning process.

Designing systematically and methodically the teaching process, based on the virtual campus is very advantageous because: it supports the training centered on learning; maintains a real, efficient and attractive training; sustains the communication between designers, teachers, computer specialists and users; facilitates diffusion and dissemination of teaching knowledge; offers practical and acceptable solutions to problems that occur during the training process; the phase of analysis sustains future development of support courses; ensures that what has been thought is necessary for reaching the learning objectives of the students and also, facilitates a correct and precise evaluation of the training process.

In terms of information resources, the virtual campus in composed of:

- data base, usually a relational data base which allows data recording and also its monitoring, accessing and documenting;
- integrated software space, which includes small codes or micro software tools;
- dynamic presentation environment, formatted data presentation through interrogation in a detailed, vivid and descriptive manner;
- audio and video environment for data transmission.

A great importance has been given to data formality, conducted by the IT team from University Henri Poincaré, after consulting the other members of the partnership. Also, it has been discussed how to place on the web (responsible Samuel Nowakowski from U.H.P.) and how to ensure the internationalization (responsible Gheorghe Popescu from U.P.B.) of the teaching resources.

The e-learning platform implemented in the virtual campus allows to: publish support courses as tutorials, case studies and links to other sources; practical exercises necessary for the professional development of an engineer by integrated experimental techniques; linear testing or self-adaptive. Also, the publishing support courses involves document publishing based on a system that implements verification and control modules and more than one version of the same support course. This fact becomes mandatory when implementing a way of working based on collaboration.

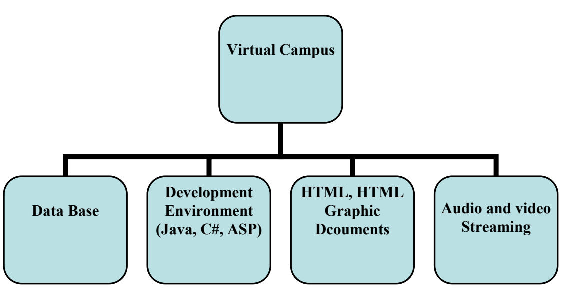


Fig. 1. Components of the virtual campus

The main components of the virtual campus are:

- *Author systems* – used for designing multimedia applications. Author systems represent high level software tools, which combine in a single package the interfaces required by multimedia design. These systems are used for designing an important content through a set of integration software tools. As a result, these products are easily used by people without previous programming experience. In EURO-QLIO project the software that has been used is „Scenari Scene".

- *Graphics*. Most of the Author systems have a combination of possibilities to import essential graphic elements, drawings and graphic files. Sometimes are available a set of animation elements, but due to the fact that the number of available features and the quality of the output varies very much function of the differences between the number of pixels of vectorial drawing functions, smoothness, contrast and color levels. Author systems offer o small number of predefined graphic features which increases lowers the amount of time required for image and graphic editing. Predefine object include rectangles, polygons, ellipses, circles and lines.

- *Animation.* This feature is a very important component of Author systems. Animation and sound are substitutes for video movement elements that are more expensive, harder to produce and bigger resource consumers.

- *Audio features.* Multimedia is often associated with graphics, animation, video effects and audio components. Audio elements are very important for a multimedia presentation, requiring lower space on hardware and lower memory than video elements. Most of the Author systems can handle audio channels separately, thus being more flexible during presentations.

- *Access policies.* In case that several persons are present at the same time, there are a fee rules that have to be followed, such as: persons should not write simultaneously and should not write on an already written area etc. These rules are also available for groups and in this case a system of monitoring different versions is required.

- *Common programs*. Sometimes the members of the same team have to work using the same program, on different computers or on the same computer (in case of programs found on a server). The application must post the same information or different on the monitor of each participant and to ensure the simultaneous access to the program commands.

- *Audio-video distribution.* The streaming delivers fluxes of audio-video data from a source (server) to several receptors (clients). The data transmissions are passive because either the users or the receptors cannot interfere into data transmission. Some applications allow a low level of interaction through a return narrow band channel.

III. CONCLUSIONS

The project „Filière Euro Qualité Logistique des Organisations", EURO-QLIO, aims to achieve a platform for distance learning in the field of Metrology, Quality Control and Organization Logistics. This platform is based on two very important components, namely: establishment of a virtual campus, on the one hand, and on the other hand, issuance of a diploma with triple recognition.

Virtual campus relies on the use of means from the Information and Electronic Communications Technology (IECT), providing a set of software tools specific to distance learning – integrated multimedia modules used for a joint education, involving both individual and group collaboration.

The diploma with triple recognition is given to the students after they graduate the learning modules coordinated from distance by the teaching teams belonging to each of the partner countries, France, Romania and Bulgaria. The diplomas obtained after going through the different learning stages are recognized by each partner due to the fact that the particularities of the national teaching system are eliminated by introducing virtual mobility both, for students and teachers.

A very important feature of this project is that it involves joint learning. Through joint learning both classical and modern ways of teaching are used. "Face to face" participation meets distance learning, each of them having their particular role.

REFERENCES

[1] A. Adăscăliței, Contributions regarding improvement of multimedia systems used at pedagogical process of knowledge asimilation in etectotechnics, *Universitatea Tehnică "Gh. Asachi" Iaşi, Facultatea de Automatică şi Calculatoare*, 2000.

[2] L. Harasim (1990) "Online education: An Environment for Collaboration and Intellectual Amplification." *In Online education: Perspectives on a new environment, p39-64*, L. Harasim ed., New York, Praeger. 7LT LB 1028.5 O58 1989.

[3] W. I. Grosky, R. Jain and R. Mehrotra (Eds.): Handbook of Multimedia Information Management. *Prentice-Hall,* 1997, ISBN 0-13-207325-0

Design, Development and Implemememtation of a Steering Controller Box for an Automatic Agricultural Tractor Guidance system, Using Fuzzy Logic

A. Carrera_González, S. Alonso_García and J. Gómez_Gil
Department of Signal Theory, Communications and Telematic Engineering
University of Valladolid, Spain

Abstract- **This study describes the development of a box for controlling the direction of an agricultural tractor, which means the movement of the steering wheel will get the desired orientation, along with its design and implementation. The final task of the box will be part of an autonomous guidance system. There will be an explanation for both software and hardware level. In addition, a brief will explain the control method used in the application: fuzzy logic. Finally, we present the experimental results obtained with step-type entries with different amplitudes, with particular emphasis on test signals similar to those obtained in the self-guided tractor that most of the time is in a straight line. Final results were satisfactory to obtain the response of the controller that is similar to the reference signal to be pursued, with a maximum delay of 2 seconds to complete the shift of direction, while for small jumps of 1.5 º is achieved the establishment in an average time of about 0.5 seconds.**

I. Introduction

The issue that occupies this study, the control theory, is an interdisciplinary field of mathematics and engineering, where it is and analyzes the behavior of dynamic systems. The objective of control is to alter the behavior of a system to present this particular conduct. At the exit you want to get the system referred to as a reference. Thus, when one or more of the variables of the output needed to keep a certain reference, a controller handle, in the right way, the plant's system to get the effect you want in the output. In [1] provides a comprehensive review of the control theory, reviewing the origins and key domains of application.

There are numerous publications in the field of control theory to illustrate in a more or less didactic, as the case, the particularities of this mode field. The book [2] provides a good reference, as it presents complex treatments of the analysis and design of control systems. In addition, it requires a comprehensive mathematical knowledge for the management of differential equations and calculating matrix mathematics among other disciplines. As far as the drivers using fuzzy logic, there are numerous publications [3], [4] and [5]. Reference [3], is a simple analysis of didactic manner, without going into unnecessary complex mathematical analysis. References [4] and [5] present a theoretical vision of such drivers supplemented with practical examples of control. There are books such as [6] where there are different components for the control systems or [7] which describes the functioning of the observers in that area.

There have been various studies for motor control applications with different techniques. To make the driver, in many cases it is necessary to identify a model for the engine. Regarding the attitude control engine using fuzzy logic are numerous studies of similar applications to which it intends to develop. Work of this kind can be seen at [8], [9], [10] and [11]. In addition, the method of control to be used in this document, there are others that have been studied and developed in various scientific papers, highlighting the classic control and easier PID. Among these, there are items like [12] and [13] where he also studied numerous ways to improve the operation. There are, likewise, many jobs where we compare the results produced by different forms of development of a single driver [14] and [15].

In regard to the implementation of final control developed, precision agriculture, has suffered a major breakthrough in recent years. There has been a revolution in terms of guidance, that has improved the performance [16]. References [17] and [18] review the state's guidance in the regions of Europe and North America respectively. There are many ways of conducting guided by artificial vision [19], using GPS and differential ways to improve the accuracy [20] and [21], using a steering geomagnetic sensor (INS) [22] or through a combination of them [23]. For the direct control of the steering actuators are used, which may consist of an electrohydraulic system [24], or simply by driving a motor connected through a pulley system that rotates in solidarity with him.

In this paper we explain how it has been designed and implemented a controller module from the direction of an

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_30,

agricultural tractor acting directly on the steering wheel. The design is divided into two distinct parts, one implements the hardware and the other part the software. To complete the study, concrete results will be presented for one of the theories of control implemented in the box. In this way, section II presents the objectives of the study and III is a brief description of the materials and real system to be monitored. In successive paragraphs IV and V will explain the hardware and software, respectively. Finally, VI presents a series of results obtained during testing validation of the box controller and the control method implemented.

II. OBJECTIVES

The aim of this study is to construct a box controller to move the steering wheel and, consequently, the address of an agricultural tractor. Thus, a design for both software and hardware levels is made in order to create a versatile device that can act as a controller with different control algorithms. Regarding to hardware, it is necessary for the system to be robust and reliable.

Once built the controller, an algorithm using fuzzy logic control will be implemented to control the positioning of the direction of the vehicle. The control through fuzzy logic is the relocation of human language to the level of control, namely to build a controller that functions intelligently as you would a person. Thus, it will make the guided according to the angular distance to be travelled to the destination point, using different functions diffuse, with triangular and trapezoidal shapes.

Once the controller box is designed, with the controller tuned properly, the results obtained on the real system will presented. To evaluate the operation a sign of stairs with different amplitudes will be used in order to have a broad knowledge of the response.

III. MATERIAL

For the tests a tractor model KUBOTA 6950DT with 53 KW of power maximum, figure 1, will be used. On this vehicle was implemented a system of steering control. The control system of management is based on a dynamo who serves as the actuator for the steering wheel through a system of pulleys and a sensor in the direction potentiometer.

The actuator system can be seen in the left of the figure 2. The advantage of using a dynamo is that by having a system operating under is not necessary to use any type of gear reducer for the movement of the steering wheel. As the dynamo interface will be used a commercial driver MD03 of the company Devantech Ltd, which provides a maximum of 20 A to the actuator of the steering wheel of the vehicle.

Figure 1: Kubota tractor 6950DT.

The direction sensor, a pot of low cost, stood at the cross of the direction of the tractor. Thus, we get a spin-axis sensor solidarity with the movement of the wheels of the vehicle. In the right of figure 2 may be the position of the pot near the left front wheel of the tractor.

Figure 2: Parts of the system of control over the direction.

For the findings of the functioning of the application in the box controller was used a laptop with an Intel Core 2 Duo 2 GHz, 1 GB DDR2 and 160 GB hard drive. The operating system used was Windows XP and the development environment on which is scheduled implementation of the computer was LabWindows CVI.

IV. HARDWARE DESIGN

For the hardware in the box controller two separate circuits are used. The first is, so to speak, the brain of the system. It is the microcontroller that is responsible for the finality of the system. The power to be used for this part of the circuit, comes from the computer which is responsible for sending the orders to the controller box. The second circuit will be an assistant to isolate the power part that connects the engine to control part connected to the microprocessor, to be funded by the tractor's battery. The voltage from the battery of the tractor may not be uniform given its nature and be connected to a DC motor that can generate spikes harmful to sensitive parts of the circuit. This will seek to avoid possible damage that might occur in the microcontroller or on the computer. In figure 3 may be a general outline of the different parts of the hardware that is the system that will be described in subsequent paragraphs.

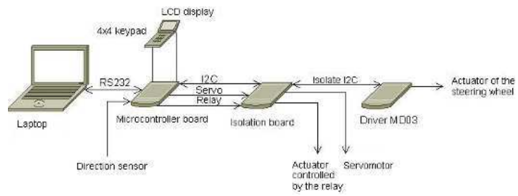


Figure 3: Outline of the system hardware

A. Microcontroller board

This board, as indicated above will be part of the intelligent controller box. It is the fundamental part of the control system: the microcontroller where the algorithm that will be responsible for the leadership position in the angle requested. Microcontroller was chosen as the PIC16F876A of the company Microchip. This device is useful because it has different interfaces for the control of different parts of the system.

The main interfaces of the system will be a 16-character LCD display and two lines and a 4x4 keypad, which will be used to display information on the application and to act directly on the system. It is also necessary to communicate with different devices in the system. For liaison with the computer will use the RS232 interface, while for sending the orders to the stage of engine power will be used to address the I2C protocol. Both interfaces are present in the microcontroller, which will facilitate, to a large extent, the communications to be carried out. For the RS232 interface is a need for a commutation of external levels for the system to work properly. This will be achieved with the help of an integrated circuit business called MAX232, which is capable of converting voltages with an RS232 to TTL Power 0 and 5 V.

In addition to the interfaces of the system, both for communication and information to the user, this board will connect the signal from the sensor to directly address one of the analog inputs of the microcontroller. It will also be necessary to send signals to the sensor power supply so the connection to this will be done through three cables. Will be generated in this part of the circuit control signal for a servomotor and a relay for the control of other tractor components such as the accelerator or the raising of the apero being processed.

As indicated above, the 5 V power necessary for the functioning of the entire plate is obtained from the computer which is responsible for sending the orders of positioning system. This is possible because the total consumption of the entire board is very small.

In figure 4 may be a comprehensive outline of the circuit built in the board of the microcontroller.

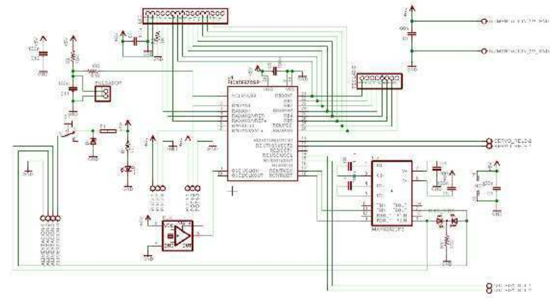


Figure 4: Designed for circuit-board PIC

B. Isolation board

The circuit part for the isolation of signals will be very important because the power system of the actuator is made from the battery of the tractor and is a source is quite irregular. Also, using a DC motor that makes the signal power of the engine from being disrupted, such as voltage spikes, which can be very damaging to sensitive parts of the module controller of the address.

To separate electrical signals are used two different magnetic insulators in the series iCoupler of Analog Devices. It will be necessary to use two different kinds of insulators. The first, it will be a special device for the isolation of signals from the I2C protocol. As a result, it has two channels of bi-directional communications. These signals are those connected with the phase of Microcontroller power. The second isolator is necessary to avoid direct connection between the signals from the microprocessor and different hardware devices in the system fed again with the battery of the tractor such as the relay and the servomotor. This second device, has two unidirectional channels of isolation perfect for the application that aims to develop, where the aim is to isolate two signals separate. One advantage of these insulators is that any external packaging through discrete components is needed, resulting in a high space savings and simplicity in the design.

Insulators require employees to use two power supplies 5 V to achieve these feeds will be used on one side of the block from the computer's microprocessor and on the other hand, the signal from the battery of the tractor stabilized by two stabilizers tension of the series 78XX properly connected through an auxiliary circuit filter.

In Figure 8 6 shows the design circuital used for the plate insulators needed in the controller box.

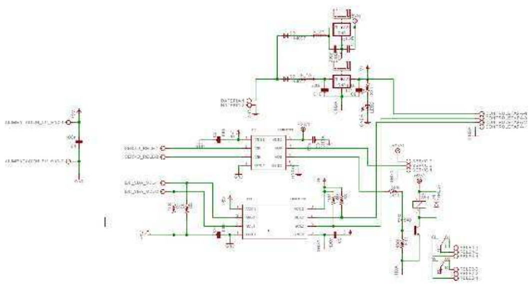


Figure 5: Circuit designed for the insulators board.

C. Power stage

To adapt the signal from the microcontroller and the board of insulation DC motor that moves the steering wheel will be necessary to add a power stage. For this purpose, was chosen the driver of MD03 commercial power of the company Devantech Ltd which will provide some signs of feeding more than sufficient for operating the engine to be monitored. Its configuration is based on a bridge in H and is capable of controlling motors up to 50 V with an intensity of 20 A. In addition, this driver allows different forms of control such as analog signals at the entrance or the one used in this case, through the I2C protocol. In Figure 6 shows the overview of the MD03 module to be used.

Figure 6: Overview of driver MD03

V. SOFTWARE DESIGN

As far as the software programmed into the microcontroller C was used as a programming language, and the C language compiler for PIC microcontrollers PCWH of the company CCS (Custom Computer Services Incorporated). Different laws will be used to control the movement of the direction, but they all use the same flow diagram which is summarized in figure 7. It may be that after an initialization of variables and the complete configuration of all devices and extras that will be used in the integrated circuit that is responsible for supervising the application, asks the user to perform a manual calibration, consisting of recording the values of the leadership on both sides, and finally in the center, or, conversely, use a calibration pre-stored on the computer. To make the distinction between whether or not it has been calibrated, the computer will analyze the message received. When it is empty, it will respond with the correct calibration message. By contrary, the computer saves the new calibration data. After the calibration of the system, the microcontroller will await the rest frame of driver settings from the computer, you'll need different kinds of parameters depending on the type of controller implemented.

When everything is configured correctly is time to come into the loop controller. The value driver will be updated with a sampling period of 10 ms by a temporary interruption. The first step of the process is to read the position, in that instant, sensing direction. Once this is done it will update the value on the LCD screen informing the user and also, if the option is active sampling external and has defeated the corresponding timer, 50 ms, a message will be sent to the computer with information from the position for the subsequent processing of the dynamics of the system. After shipments of information by various interfaces, is the calculation of the corresponding control algorithm for the configuration data that have been sent. Once the calculation is processed properly and, finally, it sends the information to the phase of power that will attack the DC motor. At this point ending the cycle of the program until the next break, at which it begins, again, the whole process of control.

First, it will receive and store the message completely, after completion of the receipt, checking the header and type of message. If both are correct, it will process data containing the message and all the variables will be updated as necessary. When you're done correctly, or the receipt, whether on the contrary, errors were detected in the header or the type of message is left of the interruption of receipt of fabric.

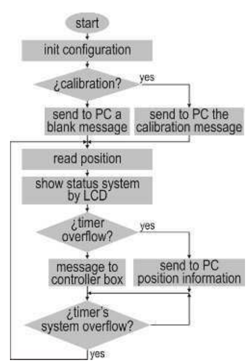


Figure 7: General organizational chart of the applications developed

To establish communication with the computer and, as might be assumed by reading the preceding paragraphs, has designed a communications system based on messages of two lengths. The different lengths are for fixed two-way communication microcontroller-computer and vice versa.

As a security measure a control of connectivity with the computer is enabled; by default, it will not be activated.

To get a complete and optimal functioning of the system were designed frames to modify all the parameters relating to the control method and the different options that allows the program.

D. Control Law implemented: fuzzy logic

For the control method that is developed will be implemented three and two triangular-shaped trapezoid, as shown in figure 8. The system of duties will focus vague on the point you want to reach and, thus, will be approaching it with a gain increasingly marginalized.

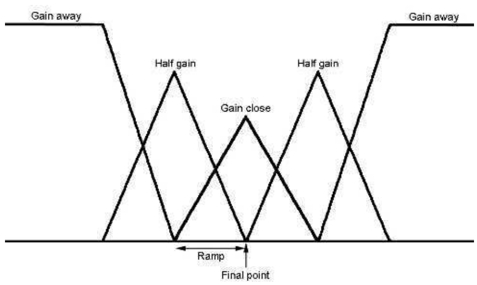


Figure 8: Functions of fuzzy logic.

For the application of control, firstly, we need to know the region of figure 8 this positioning sensor, once it knows this, apply the equation (1), where PA denotes the point where the system at that instant, O1 and O2 are the origins of the respective lines, gain1 and gain2 are the heights of each of the two lines, and finally ramp is the value of the width, as shown in figure 8, if the item is in a region with two ramps or (2) if it is on the step.

$$u(t) = \pm\left(\left(PA - O_1\right)\frac{Gain_1}{Ramp} + \left(PA - O_2\right)\frac{Gain_2}{Ramp} \right) \qquad (1)$$

$$u(t) = \pm(Gain\ \ away) \qquad (2)$$

The controller tuning is done on an experimental basis in the real system to monitor the management response to changes made in the final set of turn. The final values of the parameters used in the controller, the gains for all functions and value of the width of the ramp, are presented in (3). The value for the ramp angle is measured in the direction, that is, denotes the distance to the end point from which to begin to act that role.

$$\begin{aligned} &Gain\ \ close = 130 \\ &Half\ \ gain = 160 \\ &Gain\ \ away = 243 \\ &Ramp = 17.5^{\circ} \end{aligned} \qquad (3)$$

VI. Results

Once constructed the box for controlling the system of management were conducted numerous tests, based mainly in response to different step for changes in the angle of the wheels, to know the performance of each of the theories of control with implemented. Here are the results obtained in the case of fuzzy logic which yielded the best performance of the system. The tests were conducted for changes in the slogan of position of entry. In figure 9 the data obtained is presented graphically. The horizontal axis represents time in seconds and the vertical axis positioning units in the angle of the front of the address farming tractor to be monitored.

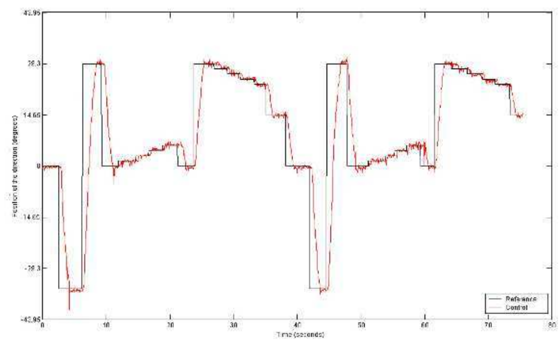


Figure 9: Results obtained for the control through fuzzy logic..

As shown in figure 9 alternated big jumps in the position of the leadership of approximately 29 degrees, with small intervals (1.46 degrees) as it would be an application of self-guided. In figure 9 can be seen that the driver still follows pretty well all the position changes that have been made, that are even smaller. The response time is slightly less than 0.5 seconds, for the worst case measured, since the microcontroller receives the new position until it begins to rotate the leadership to get to it. In addition, there is no excess, while the time needed to make a turn almost entirely on location is, from reaching the new position, slightly more than 2 seconds until the stabilization in the position of destination, although such changes are not used in applications normal guidance. Delays for small changes in the direction of the wheels of 1.5º are approximately 500 ms. Changes to intermediate values of stabilization are around 1 second as shown in figure 9. That the oscillations were observed in the area where the system is stabilized are due mainly to the quality of the sensor and in any case not exceed the value of 1º, therefore poses no obstacle for the application that aims to develop.

VII. Conclusions

In carrying out the experimental evidence has been necessary to implement a system of guided self on a real tractor. This study summed up the task of building a box for controlling this system of guided tractor.

The hardware consists of three distinct parts, on the one hand, the stage of the microprocessor fed with the signal from the computer, a second part of electrical insulators to be responsible for isolating the signals that are at the computer with the voltage coming from the battery of the tractor. Finally, phase three will be responsible for directly attacking the engine management, is the amp by the commercial driver MD03.

Regarding the software various control techniques have been implemented getting satisfactory answers in the positioning of the wheels, making the measures of this position through a sensor on the front axle of the vehicle.

The program's structure consists of a boot, a calibrated and finally the direct application of the algorithm used to control. In addition, it has enabled a system for sending and receiving frames, via post series, to achieve the fullest possible communication with the computer. With these messages all the factors of the driver will be monitored. It has also implemented a system for monitoring the physical connection to the computer in case of detecting loss and generates an emergency stop the driver.

The program that will control assistant in the computer has been developed in the LabWindows CVI, a development tool based on a similar LabView, but applying programming in C language for control. It is allowed to use a rapid development from numerous libraries and graphical environments implemented.

As evidence of performance appraisal system is a sequence of job positions that reflect the behavior of the system in a comprehensive manner, making big changes in the angle of the steering and small jumps of about 1.5 º to give idea of the behavior of applications in guided actual conventional farming. Values are obtained for the establishment of 2 seconds to complete turns of direction, but such changes are rarely carried out in guided applications. With regard to small changes in direction, of 1.5 degrees, are obtained delays of 500 ms, while for intermediate values, around 15 degrees, the values were established around a second. The results show that the behavior of the system is perfectly correct and may be used in applications where self-guided changes in the direction of the tractor will be very small as it has been studied in tests done with the real system.

VIII. References

[1] E. Fernández-Cara and E. Zuazua. "Control Theory: History, mathematical achievements and perspectives". Journal Systems Analysis Modelling Simulation 26, 2003.

[2] K. Ogata. "Ingeniería de Control Moderna" 3ª Edición. Editorial Prentice Hall, 1998.

[3] Leonid Reznik. "Fuzzy controllers", Newnes, 1997

[4] M.N. Cirstea, A. Dinu, M. McCormick, "Neural and Fuzzy Logic Control of Drives and Power Systems", Newnes Elsevier Science, Jordan Hill, Oxford, 2002

[5] F. Martin McNeill and Ellen Thro. "Fuzzy Logic: A Practical Approach", Morgan Kaufmann, 1994

[6] Christopher T. Killian. "Modern Control Technology: Components and Systems", 2ª Edición. Editorial Delmar Thomson Learning, 2000.

[7] George Ellis. "Observers in control systems", Academic Press, 2002

[8] S. Kouro R. y R. Musalem M. "Control mediante lógica difusa".

[9] Chen, T. C., Hsu, J. U., "A fuzzy Sliding Mode Controller for Induction Motor Position Control", IECON'94 20th International Conference on Industrial Electronics, Control and Instrumentation, 1994.

[10] Guor-Rurng, Chao-Lung Chiang, Ching-Tzong Su and Hong-Rong Hwung, "An induction motor position controller optimally designed with fuzzy phase-plane control and genetic algorithms", Electric Power Systems Research, 2004.

[11] Microchip Technology Inc. "Air flow control using fuzzy logic". 1997.

[12] A. Visioli and A. Piazzi. "On the Use of Dynamic Inversion for the Improvement of PID Control". IEEE/ASME Transactions on Mechatronics, 2000.

[13] J. R. Huey. "The Intelligent Combination of Input Shaping and PID Feedback control". Thesis, Atlanta, GA: Georgia Institute of Technology, 2006.

[14] Delibasi, A., Turker, T., Cansever, G., "Real-Time DC Motor Position Control by Fuzzy Logic and PID controllers using Labview", International IEEE conference on Mechatronics & Robotics '04, Aachen,Germany, September 13-15,2004

[15] Zoran Jovanovic and Goran Golo, "DC motor position control by discrete-time variable structure controllers", *Facta Universitatis*, Series Mechanics, Automatic Control and Robotics, 1997.

[16] J. N. Wilson. "Guidance of agricultural vehicles: A historical perspective". Computers and Electronics in Agriculture 25, 2000

[17] R. Keicher and H Seufert. "Automatic Guidance for agricultural vehicles in Europe". Computers and. Electronics in Agriculture 25, 2000.

[18] J. F. Reid, Q. Zhang, N. Noguchi and M. Dickson. "Agricultural Automatic Guidance Research in North America". Computers and. Electronics in Agriculture 25, 2000.

[19] C. Debain, T. Chateau, M. Berducat, P. Martinet and P. Benton. "A Guidance Assistance System for Agricultural Vehicles". Computers and Electronics in Agriculture; 2000.

[20] T. Bell. "Automatic tractor guidance using carrier-phase differential GPS". Computers and Electronics in Agriculture, 2000.

[21] T. S. Stombaugh, E. R. Benson and J. W. Hummel, "Automatic Guidance of Agricultural Vehicles at High Field Speeds", ASAE, 1998.

[22] E. R. Benson, T. S. Stombaugh, N. Noguchi, J. D. Will and J. F. Reid. "An Evaluation of a Geomagnetic Direction Sensor for Vehicle Guidance in Precision Agriculture Applications", ASAE, 1998.

[23] K. Moriwaki. "Autonomous Steering Control for Electric Vehicles Using Nonlinear State Feedback H_∞ Control". Article Nonlinear Analysis, 2005.

[24] Q. Zhang, D. Wu, J.F. Reid and E. R. Benson, "Model recognition and validation for an off-road vehicle electrohydraulic steering controller", Mechatronics,vol. 12, 2002.

A Screen Sharing Broadcast Scheme in Distance Learning

Tzong-An Su and Han-Chun Young
Department of Information Engineering and Computer Science
Feng Chia University
Taichung, Taiwan, Republic of China
tasu@fcu.edu.tw

***Abstract*— VNC is popular and stable open source remote control software. It is widely used in many applications, such as the screen sharing part of a distance learning supporting system. VNC is designed originally for one-to-one remote control application, which means that the bandwidth constraint is not seriously considered. In many distance-learning applications, the screen image is broadcasted to many students, thus the performance become degraded. This paper focuses on this problem, and proposes an algorithm to improve the screen sharing broadcast scheme on VNC systems and makes it more suitable for distance-learning applications.**

I. INTRODUCTION

E-learning plays an important role in the IT industry. One of the most researched topics in e-learning area is distance-learning. Distance learning is a field of education and a part of e-learning that uses various technologies and instructional systems to deliver education to students who are not physically "on site". In distance learning, teachers and students may communicate asynchronously/synchronously by exchanging teaching material, vocal discussion, instant messages, files, etc. For example, in [1], we have integrated several open source software systems like Moodle, DSpace, and the Asterisk VoIP network together to develop a virtual learning environment, called Distant VCR, to support the need of distance learning.

A distance learning supporting system is an integrated technology, composed of learning management, voice over IP, screen sharing, etc. Screen sharing is a critical part in such a system. It is used to deliver the visual teaching material from the teacher to students. For example, whatever shown on the teacher's PC screen such as Power Point files, pictures, papers and videos, can be synchronously shown on students' PC screens.

Due to the fast evolution of computer hardware and networking technology, screen sharing technology is continuously developed and improved by many companies. Among those Screen Sharing technologies, VNC (Virtual Network Computing) is popular and stable open source software. It is widely used in many applications, such as the screen sharing part of the web phone system [9], remote computer system diagnosis, e-learning, etc.

VNC is designed originally for one-to-one remote control application, which means the bandwidth requirement is not seriously constrained. When using in distance-learning applications it has to broadcast the teacher's screen image to many students' computers, thus the performance become degraded. This paper focuses on this problem, and proposes a broadcast scheme enhanced from [4] to improve screen sharing broadcast on VNC system and make it more suitable for distance-learning applications.

The structure of this paper is composed as follow: in Section 2 we introduce the VNC system. Our solution of screen sharing broadcasting is presented in Section 3, and the experiment/simulation result is shown in Section 4. Finally, Section 5 is the conclusion.

II. VIRTUAL NETWORK COMPUTING (VNC)

NC stands for Network Computing, is a remote control technology which users can use simple, cheap devices to access central resources. VNC (Virtual Network Computing) [8] is a remote control system based on RFB (Remote Frame Buffer) [10] protocol, derived in 1998 by UK Cambridge University and AT&T Co. VNC is a thin-client system, main overload is on the server. Therefore the implementation of client is easy and cheap. The client can run on mobile or laptop device.

The architecture of a VNC system is shown as Fig. 1. The system is composed of two main part, vnc server and vnc viewer (client) with RFB (Remote Frame Buffer) protocol running between them. The VNC server is composed of screen capture and remote control two main parts. The screen capture part monitors the changes of screen and saves them into a memory buffer. When a viewer sends a screen update request, the server compares the buffered data, finds out the change parts of screen, compresses and sends the screen update data to the viewer. The remote control part receives control messages from viewers and completes the task running on the server according to control messages. The VNC viewer is also composed of two main parts, screen display and remote control. The screen display part receives the screen update data from the VNC server by RFB protocol and display them on the viewer. The remote control part sends the control message to the server to complete the task.

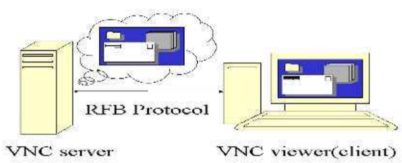


Fig.1. Architecture of VNC System

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_31, © Springer Science+Business Media B.V. 2010

III. The Screen Sharing Broadcast Scheme

To solve the problem of screen sharing broadcast described in Section 1, we have surveyed some related works. Paper [6] gave an algorithm which found a best broadcast scheme from an un-weighted tree (all edges have unit weight). Paper [5] extended results from [6] and found the best broadcast scheme for a weighted tree (with different weights assigned to edges of a tree). Since to find the best broadcast scheme in a graph is NP Complete as stated in [6], TBA (Tree Based Algorithm) presented in [4] proposed an "efficient" broadcast scheme for an un-weighted graph (instead of finding the best scheme).

In a distance learning environment, computers used by students and teachers form a network. Due to the different latency of real network environment, the graph transferred from a network topology often gets assigned with edge weights to represent the latency. Thus our problem is transformed into finding an efficient broadcast scheme from a weighted graph. Fig. 2 shows a graph with weighted edges.

As shown in Fig. 2, assume node A is the teacher's computer and all other nodes represent students' computers. To broadcast teacher's screen to all students, a possible broadcast scheme might be (AB, AG, BC, BD, BF) or it might be (AB, AG, BC, BD, GF) or might be other schemes. The first scheme takes 25 at best to complete the broadcast while the second one does it in 16. Note that we have allowed parallel data transmissions in the network. Our algorithm described in this section will find an efficient way to do the screen broadcast in terms of the total time it takes.

The following sub-sections describe our algorithm which finds an efficient broadcast scheme form arbitrary root of an arbitrary weighted connected graph. We named it TBA+.

A. Definitions and Assumptions

In order to describe our algorithm, we make the following assumptions.

A1: Two nodes x and y in a network may have a link xy. We give this link an edge weight defined by the latency of the link. The edge weight is a positive integer and in direct proportion to the latency. It is a relative value, so there is no need to assign a unit.

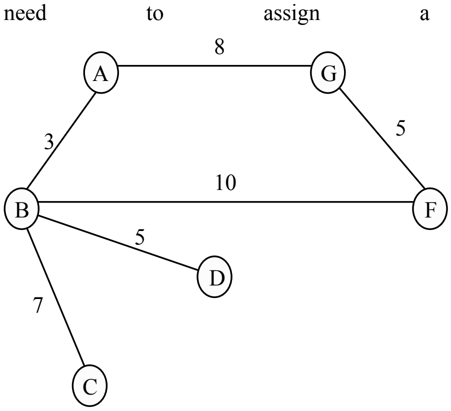


Fig.2. A graph with weighted edges

A2: Whether x connects to y or y connects to x, the link's edge weights derived from both directions are the same. The value of the edge weight is a static value.

A3: A node x can only inform (or communicate with) any adjacent node y.

A4: One node can inform one and only one node at the same time. It must wait for the end of the current communication before starting another communication.

A5: Our broadcast mode is simplex. Nodes can act either as sender or as receiver but not both at the same time.

A6: We assume the synchronization time of any two nodes before transmitting data is very small, and can be ignored.

A7: The handoff time (time between a vnc server finishes one transmission and starts the next transmission) among vnc viewers linked to one vnc server is very small, and can be ignored.

A8: The time spent on one data transmission between any two adjacent nodes is the same as the edge weight of the edge connecting these two nodes.

We also include some definitions here.

D1: Informed Node: An informed node is a node which has received the broadcasted screen contents.

D2: Uninformed Node: An uninformed node is a node which has not received the broadcasted screen message.

D3: Source Node: Teacher's PC.

D4: Listening Node: all students' PCs.

B. Introduction to TBA+ Algorithm

The algorithm is composed of three main parts, Cal_Dvt, Cal_node_weight, and Cal_match. It repeats running the three parts until all nodes is informed. In finding one match, we first process Cal_Dvt to find the depth of each node. We used a modified weighted BFS (Breadth First Search) to find the depth. Second, we process the Cal_node_weight procedure to calculate the node weight of each uninformed nodes by the Dvt value. Final, in procedure Cal_match, we find the match between the informed nodes and uninformed nodes by the node weight.

The detail procedure of TBA+ is presented in Appendix.

C. Complexity of TBA+

We will analyze the complexity of our algorithm TBA+ by finding one match in the worst case. We represent the number of all nodes in a graph as V, number of all edges as E. First, in the Cal_Dvt procedure, we reset all nodes value. This will cost O(V) time. Then we use BFS to find out the Dvt values of all uninformed nodes, In the worse case this process takes $1+2+3+\ldots+V = (1+V) \times V / 2$, about $O(V^2)$ time. Next, in the Cal_node_weight procedure, we consider the complexity of the sub-procedure Weight first. In the sub-procedure Weight we need to sort all children nodes according to the node weight and it costs O(V logV) time. We also need to find the max-weight which takes O(V) time. So, sub-procedure's complexity is O(V logV). Back to the procedure Cal_node_weight, we need to reset all node_weight first, it costs O(V). Next we use Remote queue recursively to calculate the node_weight by calling the sub-procedure Weight. So the complexity of Cal_node_weight is $O(V) \times O(V \log V) = O(V^2 \log V)$. Finally in the Cal_match procedure we need to find the match by the order of nodes' degreeU. To sort out this order costs O(V). We check all connected edges of each node by the order to find a match, cost O(V) x O(degree(V)) = O(V) x (E/V) = O(E) time. We also need modifying the degreeU value of nodes affected. This action costs O(V) x O(degree(V)) = O(V) x (E/V) = O(E) time. Therefore, Cal_match procedure's total time complexity is O(E).

From this analysis we get the time complexity of running one match of the TBA+ algorithm is $O(V^2) + O(V^2 \log V) + O(E)$. Our graph must be a simple connected graph, $V^2>E$. Thus, the total time complexity to run one match by TBA+ algorithm is $O(V^2 \log V)$.

IV. EXPERIMENTAL RESULTS

We used two experimental ways to verify our algorithm. First, we ran our algorithm in a LAN (Local Area Network) environment. Next, to show the effect of adding edge weights, we simulated our algorithm in a LAN and compare our algorithm TBA+ with the TBA algorithm in [6]. The results showed that either in LAN experiment or in Simulation experiment, our algorithm TBA+ have overall best performance.

A. LAN experiment

We ran our experiment in a simple LAN environment to verify our method. In a LAN environment, each node is connected to all other nodes. Also, the latency between any two nodes is very small and each link can be considered as having the same weight. Therefore, the topology of the network is an un-weighted complete graph. The broadcast schemes derived by TBA+ and TBA should be the same. So we only compared TBA+ with origin directed connect broadcast scheme (that is, all students' PCs were connected to a teacher's PC).

As measuring the receiving time of each packet among listening nodes is hard to achieve, we monitor the number of received bytes during fixed time period of each listening nodes instead. We calculate the number of bytes received by all listening nodes during this time interval and find the average byte receiving rate (ABRR) in terms of bytes/second. Also, we assume the best ABRR is when only one listening node connects to the source node in the LAN. We define a measurement called efficiency as follows.

$$\text{efficiency} = \text{ABRR} / \text{best ABRR} \qquad (1)$$

The result is shown below:

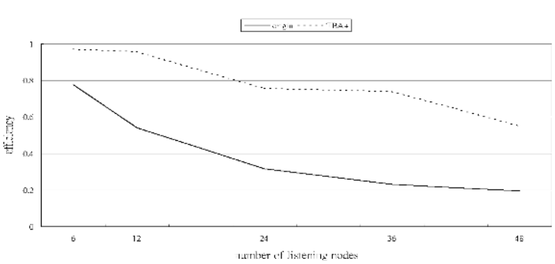


Fig.3. Efficiency in a LAN environment.

As Fig.3 shows, the performance is decreased with the growth of the number of the listening nodes. The total bandwidth is limited in the LAN. As the growth of the number of listening nodes, each listening node's bandwidth is getting smaller. Thus we find that in the LAN environment, bandwidth is more important than latency.

B. Simulation of Broadcast in WAN environment

In the distance learning application, nodes are normally located distance apart. Thus, to see the algorithm's performance, we have to run our algorithm in a simulated WAN (wide area network) environment. We modified the VNC source code by adding delay in the sending message part of the VNC viewer to simulate latency in WAN. The topology of a WAN is a weighted connected graph, so we only compare TBA+ and TBA here.

In the simulation experiment, we used 25 random weighted connected graphs as variety of network topology. These graphs are divided into 5 groups by the number of listening nodes. Each graph was given an ID as nXsY, X means the number of listening nodes, Y means the order of the group. We used the same measurement and efficiency equation (1) in the previous section to do the comparison. The result is shown below.

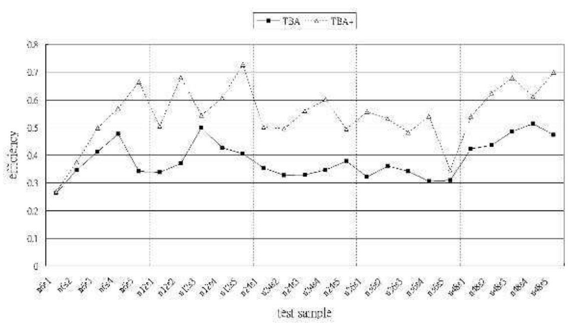


Fig.4. Efficiency in a simulated WAN environment.

From Fig. 4, the TBA+ algorithm is always better than the TBA algorithm with some graphs shown almost double efficiency. One side observation showed that the efficiency is not proportion to the number of listening nodes. Instead, we find the ability of root and the structure of a graph are some time key points to affect the efficiency of a screen sharing broadcast.

V CONCLUSION

In this paper, we propose a new algorithm to find an efficient broadcast scheme for the screen sharing broadcast in a VNC system. This new algorithm can find efficient broadcast scheme from arbitrary root in an arbitrary weighted graph. The experimental result also shows that either in a LAN environment or in a WAN environment, our method has much better performance than previous proposed one, namely TBA in [6].

The experimental results also reveal that both bandwidth and latency of the network affect the performance of a screen sharing broadcast in VNC system. In the LAN environment, the bandwidth is more important. In WAN, the latency, structure of the graph, ability of root, is more important than bandwidth.

APPENDIX

Herustic TBA plus(tree broadcast algorithm plus)
Input: arbitrary graph G=(V,E), with informed nodes u as root node.
Output: a broadcast scheme.

```
1.   round = 0; // initialize some value
2.   round_step = 0;
3.   Remote = NULL;
4.   Passed = NULL;
5.   Uninformed = |V| - 1; //number of uninformed
6.   while( Uninformed != 0 ){
          round = round + round_step
          procedure Cal_Dvt
          procedure Cal_node_weight
          procedure Cal_match
7.   } // end
```

procedure **Cal_Dvt**

```
1    Remote.clear(); // clear the queue
2    for( all nodes v ){ // do some initial , reset
     2.1   v.childrenset.clear();
     2.2   v.node_weight = 0;
     2.3   if( v.informed ){   v.Dvt = v.Dvt - round_step; }
     2.4               else{   v.Dvt = 0;   }
     2.5   v.parentN = 1;
     2.6   v.degreeU = 0;
3    }//for
4    Passed ← all informed nodes;
5    while( Passed not empty ){
     5.1   node v = Passed.dequeue() ;
     5.2   for( all uninformed neighbor nodes w of v ){
        5.2.1  if( w.Dvt == 0 ){
                   w.Dvt = v.Dvt + vw.edge_weight;
                   Remote.enqueue(w);
                   //add to Remote for Cal_weight.
                   Passed.enqueue(w);
        5.2.2  } else if (w.Dvt > v.Dvt+vw.edge_weight){
                   w.Dvt = v.Dvt+ vw.edge_weight;
               }//if
     5.3   } //for
6    } //while
7    //  end
```

procedure **Cal_node_weight**

```
1    while( Remote is not empty ){
     1.1   v = Remote.dequeue();
     1.2   if( v.childrenset == 0 ){ v.node_weight = 1; }
     1.3   else{v.node_weight = Weight(v.childrenset); }
     1.4   v.node_weightTemp = v.node_weight;
     1.5   for( all uninformed neighbor w of v ){
        1.5.1  if(   w.Dvt + wv.edge_weight == v.Dvt ){
        1.5.2      if( v is not in w.childrenset ){
        1.5.3          w.childrenset ←  v;
        1.5.4          w.degreeU += 1;
        1.5.5          v.parentN += (1/ wv.edge_weight);
                   }//if
        1.5.6      if( w not in Remote &&
                       w is uninformed){
        1.5.7          Remote.enqueue(w);
                        // add w to the end of queue for
                        // recount w.node_weight
                   }//if
               }//if
     1.6   }//for
     1.7   v.node_weight = v.node_weightTemp / v.parentN;
}// while
```

Procedure **Weight**
Input: v.childrenset
Output: the weight of node v

```
1    sort v.childrenset by children.node_weight in decrease
     order
2    edgeWeightSum = 0;
3    weight_Max = 0;
4    for( all w in v.childrenset ){
     4.1  edgeWeightSum += wv.edge_weight
     4.2  if(weight_Max<w.node_weight+edgeWeightSum){
     4.3     weight_Max=w.node_weight+edgeWeightSum;
          } //if
     }//for
5    return   weight_Max;
```

Procedure **Cal_match**

```
1    for( all informed nodes v, who have minimum degreeU,
     do once ){
     1.1  if( v.Dvt > 0 ) continue;
     1.2  if( v.degreeU <= 0 ) continue;
     1.3  MaxNode = v; MaxEdgeWeight = 0;
     1.4  for ( all uninformed neighbor nodes w of v ){
          1.4.1  if( w.Dvt >= v.Dvt + vw.edge_weight )
          1.4.2    if(w.node_weight> MaxNode.node_weight )
          1.4.3      MaxNode=w;
                     MaxEdgeWeight = vw.edge_weight;
          }//for
     1.5  if( MaxNode == v ) continue; // v have no match
     1.6  MaxNode.informed=ture;
          Uninformed = Uninformed - 1;
     1.7  MaxNode.Dvt = v.Dvt + MaxEdgeWeight;
          v.Dvt = MaxNode.Dvt;
     1.8  for( all informed neighbor z of MaxNode ){
          1.8.1   z.degreeU = z.degreeU - 1;
          }//for
     1.9  vw.sendToNode = MaxNode;
     1.10 vw.sendAtRound = round;
     }//for
2    round_step = min{ v.Dvt },
     v is all informed nodes with v.Dvt > 0;
```

REFERENCES

[1] T. A. Su, “Distant VCR – A Virtual Classroom for Distance Learning”, First IEEE International Symposium on Information Technologies and Applications in Education, 2007. ISITAE apos;07. Volume , Issue , 23-25 Nov. 2007 Page(s):174 - 178

[2] Albert M. Lai, Jason Nieh; "On the performance of wide-area thin-client computing"; ACM Transactions on Computer Systems (TOCS) Volume 24 , Issue 2 (May 2006) Pages: 175- 209. 2006.

[3] Claire Kenyon, Nicolas Schabanel; "The data broadcast problem with non-uniform transmission times"; Symposium on Discrete Algorithms, Proceedings of the tenth annual ACM-SIAM symposium on Discrete algorithms, Pages: 547-556, 1999.

[4] Hovhannes A. Harutyunyan, Bin Shao; "An efficient heuristic for broadcasting in networks"; Journal of Parallel and Distributed Computing, Volume 66 Issue 1 (January 2006) Pages: 68-76. 2006.

[5] Jae-moon Koh, Dong-wan Tcha; "Information Dissemination in Trees with Nonuniform Edge Transmission Times"; IEEE Transactions on Computers Volume 40, Issue 10 (October 1991), Pages: 1174 - 1177, 1991.

[6] P.J. Slater, E.J. Cockayne, S.T. Heditniemi; "Information dissemination in trees"; SIAM J. Comput. 10 (4) Page:692-701, 1981.

[7] Rene Beier, Jop F. Sibeyn; "A powerful heuristic for telephone gossiping"; The Seventh International Colloquium on Structural Information & Communication Complexity (SIROCCO 2000), Pages:17-36, 2000.

[8] Richardson, T.; Stafford-Fraser, Q.; Wood, K.R.; Hopper, A.; “Virtual network computing”; Internet Computing, IEEE Volume 2, Issue 1, Jan.-Feb. 1998 Page(s):33 – 38. 1998.

[9] TelTel- More than a telephone, it's a TelTel Phone. http://www.teltel.com/tw/index.html.

[10] Tristan Richardson RealVNC Ltd; “The RFB protocol”; Version 3.8 October 2006.

Robotics-based Curriculum Development for An Immigration Course into Computer Systems Engineering

C. N. Thai[1], J. M. Mativo[2] and G. Clinton[3]
University of Georgia
[1]Biological and Agricultural Engineering Department
[2]Workforce Education, Leadership and Social Foundations Department
[3]Educational Psychology and Instructional Technology Department
Athens GA 30602-4435
E-mail: thai@engr.uga.edu
Web site: http://www.engr.uga.edu/people/thai

***Abstract* - A robotics-based freshman immigration course into Computer Systems Engineering (CSE) was created to introduce students to the main design issues in CSE: multi-controller programming, sensor interfacing, servo actuation and serial communications. The instructional materials were designed around 3 projects: car bot in maze, GERWALK bot penalty kick, and GERWALK bot going up stair steps. Selected descriptive and prescriptive creativity tools were also incorporated into the curriculum to round out student design experience. Mid-term course assessment results are presented in this article.**

I. Introduction

Starting in Fall 08, the Faculty of Engineering at The University of Georgia is offering a new Bachelor of Science degree in Computer Systems Engineering with an interdisciplinary perspective:

> "The Computer Systems Engineering undergraduate program of study covers the range from scientific principles and engineering concepts to the practical application of designing and implementing products and systems requiring hardware-software integration. Graduates from this program will have a critical understanding of computer hardware and software development issues and how to use computers to automate, monitor and control various systems, industrial as well as biological. The job market needs of a computer systems engineering graduate include telecommunications (wired and wireless networks), computer-human interactions (training or role-playing), decision-making (financial and health care industries), and embedded mechatronics (industrial or biomedical), where the engineer is needed to increase the efficiency and overall effectiveness of computing and sensing systems".

(http://www.engineering.uga.edu/academics/Comp_Sys.php)

Recognizing the importance of freshmen engineering courses in affecting students continuing interests and retention ([1],[2],[3]), we opted to create a new 1 semester-credit hour course whereas the students can obtain an overview of the main issues that a Computer Systems Engineer would encounter in his or her work after graduation. We also wanted to showcase selected creative problem-solving approaches and tools used, thus the core instructional materials were designed around 3 robotics projects:

1. The first project involved a pre-made "car-bot" equipped with NIR sensors to help it avoid obstacles using a provided software solution and the student's task was to extend this solution to allow this car-bot to navigate itself autonomously through a maze (see Fig. 1).

Figure 1. Car-bot traversing a maze.

2. The second project involved a GERWALK bot (for a definition of "GERWALK", please visit this web site http://en.wikipedia.org/wiki/VF-1_Valkyrie). Essentially, it was a "box" on top of 2 legs allowing bird-like bipedal motion (see Fig. 2).

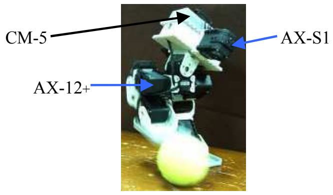


Figure 2. GERWALK bot kicking a tennis ball.

Also starting from a provided software solution, the student's task was be to extend this software solution to allow the GERWALK bot to kick a ball into a goal guarded by a car-bot controlled by the instructor.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_32,

3. The third project required the student to adapt the existing GERWALK bot hardware and software design so that it can climb up stair steps in a balanced manner (see Fig. 3).

Figure 3. GERWALK bot climbing stairs.

The following demo web sites were created for the students:

- http://www.engr.uga.edu/~thai/Bioloid/CarBotMaze/
- http://www.engr.uga.edu/~thai/Bioloid/KickBall/
- http://www.engr.uga.edu/~thai/Bioloid/GerwalkStairs_2/

The secondary objectives of the course were to introduce students to the main steps of the engineering design procedure and selected creativity practices, in preparation for their first formal course in engineering design to be taken during their sophomore year.

The course specific goals were for students to learn about the basic programming of microcontrollers, the interfacing of sensors for sound and near-infrared light, the actuation of servo motors and computer serial communications concepts. These concepts and methodologies were demonstrated "synchronously" in a technology-enhanced classroom described in [4], and also "asynchronously" using narrated tutorials pod-casted from the web when students were outside the classroom. The students were also introduced to selected descriptive creativity tools such as "Empathy" [5] and "SCAMPER" [6], and a prescriptive creative design approach called TRIZ [7].

II. Curriculum Description

A. *Hardware/Software Platform Chosen*

The main challenge in this curriculum design was to find an appropriate hardware/software platform for freshman engineering students who may not have yet been exposed to software programming. Thus we wanted a more limited software development environment than standard tools such as BlueJ (http://www.bluej.org) or Alice (http://www.alice.org), so as to lower the necessary learning curves for students. On the other hand, we wanted a hardware-rich system so as to facilitate hands-on projects and keep students' interests high early and throughout in the course. Thus we decided on using the Bioloid system from Robotis (http://www.robotis.co) which was based on Atmel's ATmega chips (http://www.atmel.com). The system consisted of a main controller CM-5, based on the ATmega128, which could be interfaced with smart actuator modules called AX-12+ and smart sensor modules called AX-S1 via TTL serial protocols (see Fig. 2). The AX-12+ and AX-S1 modules were controlled by their individual ATmega8 microprocessors. The CM-5 could be programmed to tell a given set of AX-12s to go to individual and different Goal Positions (0-1023 possible values corresponding to a physical 300^{o} range of motion) and the AX-12s could report their Present Positions, Speeds, Loads and Temperatures back to the CM-5 during their progress towards the prescribed Goal Positions. The AX-S1s could provide active or passive NIR detection and also detect sound levels (both with 0-255 possible values for sensor data range). The Bioloid system provided students with 3 PC-based software tools: Behavior Control Programmer (BCP), Motion Editor (ME) and Robot Terminal (RT) [8]. These tools communicated with the CM-5 via the PC's RS-232 communication port.

The Behavior Control Programmer tool is a context-sensitive line-editor (see Fig. 4) wherein students could create bot programs, check them for logic errors, download the compiled versions down to the CM-5 and run them from the PC or autonomously on the CM-5 as needed. To create a program line, the student would double-click on the left most cell to pop up a small window showing all the available commands such as IF, ELSE IF, ELSE, CONT IF, JUMP, LOAD, COMPUTE, START, END, CALL and RETURN. Once this command was selected, the student double-clicked on the next cell on the right which would display additional possible data entries that were context-sensitive based on the actual choice made for the previous cell, and so on (see Fig. 4).

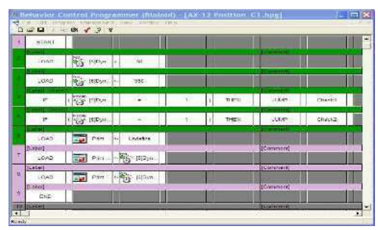


Figure 4. Windows Interface of Bioloid BCP tool.

The Motion Editor tool (see Fig. 5) was used for "motion programming", i.e. creating coordinated servo positions of a group of servos belonging to a given bot to obtain a required bot "pose". The ME tool used graphics to help students create sequential "poses" that the bot must execute at run time in order to perform a wanted overall maneuver of the bot. The ME tool had the capacity for 127 motion pages with a maximum of 7 poses in each motion page. This tool was designed around a "teach and learn" paradigm which made the motion programming task quite easy: the student would first select a group of servos to disable the power to them so that their positions could be manually adjusted to achieve a wanted pose ("teaching" phase), then the student would reactivate power to those servos so that the ME tool could record their final positions resulting from that prescribed pose ("learning" phase). Due to the execution speed of the Bioloid system, all poses had to be statically stable by them-selves to ensure successful overall maneuver execution. The BCP and ME tools were designed to be used in concert to create sophisticated and intelligent behaviors in 12 bots pre-designed by Robotis such as the GERWALK and HUMANOID models

[9]. For custom-made bots, the user would have to use the Robot Terminal tool which was beyond the scope of this report.

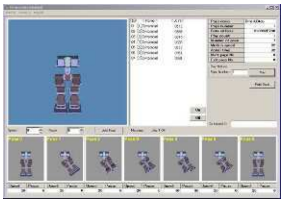


Figure 5. Windows Interface of Bioloid ME tool.

B. Instructional Technologies Used

Classroom instruction was done in a technology-enhanced teaching laboratory whereas each student had access to a Windows Vista PC equipped with dual displays to achieve a dual-workspace capability for in-class activities [4]. Students used one display for their own personal workspace, and the second display was configured as an "external" workspace which could be set to be "passive" and only received live transmission of all activities currently happening on the desktop of the Teacher Station (such as PowerPoint slides). Students could then optionally use their Pen Tablets to capture and annotate screen shots of this external workspace and thus could create their own electronic class notes as separate graphics files or integrate them into their own Word or OneNote documents to take home. A second option was also provided to students by the instructor's pre-printing all PowerPoint slides (or any other Office 2007 documents) into a Windows Journal Writer file format which essentially transformed these slides into graphical backgrounds un-removable by the students but that can be annotated "on-top" by students using Pen Tablets (please note that this option is only available to those using Office 2007 with Windows Vista or Windows XP Tablet PC for operating systems). However this "external" workspace could also be used in an "active" mode allowing students to actually share into the desktop of the Teacher Station or peer student's desktop to participate in collaborative works (see [4] for more details).

This teaching laboratory was also equipped with microphones throughout the room so that the instructor's lecture narrations and verbal exchanges among instructor and students, along with instructor's desktop activities, were recorded using Techsmith's Camtasia Studio software and then processed off-line to be published on the UGA WebCT facility within 1-2 hours after the end of each class session.

The Camtasia software was also used to create custom multimedia presentations that were available 24/7 from the web (via WebCT) and served either as lead-in components to an upcoming lecture (such as showing students how to use the BCP tool, or explaining a pre-made Bioloid program that was to be used as the starting point for the next homework), or as homework feedback to clarify some discovered student misconceptions and presenting a more efficient solution, etc… The goal was to connect in-class and outside-of-class activities into a chain of learning episodes, all aimed towards student learning of the subject matter in as many modes as possible, and as often as students would want it.

C. Pedagogical Approach in Bioloid Programming Tasks

Our overall approach was to consider "teaching as fostering learning" [10] or at least to make students aware of their strengths and weaknesses in the subject matter.

For the Fall 08 semester, this course was scheduled for one 55-minute meeting weekly and started out with the instructor going over the syllabus particulars and doing a "lab" showing students how to use the dual-workspace capabilities that they had at their disposal and enrolled students were requested to take several asynchronous surveys:

1. A WebCT survey to provide information about the current knowledge levels of enrolled students in Computer Systems Engineering.
2. A learning style survey at the University of Arizona's site http://www.ulc.arizona.edu/learning_style.php to help students determine their dominant learning styles: visual, auditory or kinesthetic.
3. Two "creativity" surveys [11] and [12] for students to correlate their work habits with those of highly creative individuals.

The second lecture was to provide feedback over the results of the previous surveys and to lecture on general Engineering Design approaches (descriptive and prescriptive methods) and to discuss the main issues that a Computer Systems Engineer would have to consider such as: sensor interfacing, mono and multiprocessor programming, actuator control and network/multiprocessor communications.

The next series of 3 lectures were focused on using the BCP tool to program the CM-5 and the servos AX-12+ to perform the following selected tasks:

1. Program the CM-5 to declare and assign values to variables and to print outputs using the commands LOAD and COMPUTE.
2. Design selection structures to model alternate actions to undertake with IF, ELSE IF, CONT IF commands.
3. Create deterministic and conditional loops using the IF and JUMP combination and statement labels.
4. Design and use appropriately subroutines with the CALL and RETURN commands.
5. Set AX-12 Moving Speed and Goal Position.
6. Understand and use the servo Moving Flag parameter to monitor the Present Position and Present Speed of a servo while it is moving to a set Goal Position.

All lectures and assignments during this period were designed to allow the student to progress from the handling of one servo to two servos, first from a mono-processor viewpoint (i.e. CM-5 only) and then from a multi-processor viewpoint (i.e. CM-5 in a supervisory mode checking in on progress made by each AX-12's controller in achieving a certain task assigned to them).

First using only one AX-12 servo, students were asked in class to program the CM-5/AX-12 system so as to move the servo back and forth between 2 Goal Positions (800 and 200 for example), which they would respond with the following **Program 1** (in pseudo code):

- Label_1:
  LOAD Servo1 GOAL POSITION with 800;
- LOAD Servo1 GOAL POSITION with 200;
- JUMP to Label_1;

Although **Program 1** contained no logic error, when it was executed students could see that Servo1 just sat there and vibrated, then they realized that they did not take into account the fact that Servo1 needed a finite time period to actually travel to one Goal Position to the next. Thus they next inserted Wait loops based on the Moving Flag value (which is set to 1 by the AX-12's controller as long as the servo is in motion towards a given Goal Position) to yield **Program 2**:

- Label_1:
  LOAD Servo1 GOAL POSITION with 800;
- Wait_1:
  IF Servo1 MOVING FLAG is 1, THEN JUMP to Wait_1;
- LOAD Servo1 GOAL POSITION with 200;
- Wait_2:
  IF Servo1 MOVING FLAG is 1, THEN JUMP to Wait_2;
- JUMP to Label_1;

Next students were asked to expand **Program 2** into a weekly homework assignment whereas 2 servos were to be moved between different Goal Positions (for example between 1000 and 300 for Servo2) and at different speeds (one high, set at 1000 and one slow, set at 200). They most invariably came up with a solution similar to the following **Program 3**:

- Set Servo1 MOVING SPEED to 1000;
- Set Servo2 MOVING SPEED to 200;
- Label_1:
  LOAD Servo1 GOAL POSITION with 800;
- LOAD Servo2 GOAL POSITION with 1000;
- Wait_1_1:
  IF Servo1 MOVING FLAG is 1, THEN JUMP to Wait_1_1;
- Wait_1_2:
  IF Servo2 MOVING FLAG is 1, THEN JUMP to Wait_1_2;
- LOAD Servo1 GOAL POSITION with 200;
- LOAD Servo2 GOAL POSITION with 300;
- Wait_2_1:
  IF Servo1 MOVING FLAG is 1, THEN JUMP to Wait_2_1;
- Wait_2_2:
  IF Servo2 MOVING FLAG is 1, THEN JUMP to Wait_2_2;
- JUMP to Label_1;

Students would observe that **Program 3** did move the 2 servos back and forth as required, however they also noticed that the faster servo was waiting on the slower one to get to its intended Goal Position before they continued their motion cycles all over again. Here, when asked how to make the servos' motions independent of each other, most students would not be able to come up with a workable solution. Thus the instructor next created a homework-feedback narrated tutorial showing first a video clip showing that the 2 servos could actually be programmed to move back and forth independently of each other, and explained that the "wait-and-see" attitude of the faster servo was due to the use of the previous Wait loops, and therefore the challenge was how to wait for the servos to get to their intended Goal Positions without using the Wait loops, which seemingly was a contradiction of ideas. This narrated tutorial next provided some hints about the "correct" solution by sharing with students the main section of the new **Program 4**:

- Label_1:
  IF Servo1 MOVING FLAG is NOT 1, THEN CALL Move_1;
- IF Servo2 MOVING FLAG is NOT 1, THEN CALL Move_2;
- JUMP to Label_1;

Here **Program 4** showed a 180^o turn in thinking about the problem, because now the emphasis was on what to do when the servos were found to be not moving (because they had arrived at their intended Goal Positions), instead of figuring out what to do when they were found to be moving as previously done for **Programs 1, 2 and 3**. In other words, students had to think about the parallel tasks of the other processors in the system (i.e. the ATmega8 inside each AX-12) and not just consider that only the CM-5 is doing all the "thinking" and "acting" for the whole bot assembly. The ultimate goal of these types of exercises was to convince students to be comfortable with an "Opposable Mind" [13] and to consider task scheduling when dealing with a multi-processor system such as the Bioloid system. However detailed code for subroutines Move_1 and Move_2 were not shared to students as this task became their next homework assignment.

Looking at the above programs in a different light, we could say that we introduced the students to multi-threaded programming without burdening them with too much abstraction, as a matter of fact they had learnt to associate the threads Move_1 and Move_2 with concrete objects as Servo1 and Servo2 communicating back to the CM-5 using their respective MOVING FLAG as semaphores.

Similar instructional approaches would be applied when providing student practice in interfacing and programming the sensing modules AX-S1, specifically in active NIR sensing and sound detection.

D. Introducing Students to Selected Creativity Tools

Due to lack of time, we could only spare 4 meetings to discuss and apply selected creativity tools: 2 meetings on descriptive tools such as "Empathy" [5] and "SCAMPER" [6], and 2 meetings on a prescriptive tool called TRIZ ("Teoriya Reshentya Izobretatelskikh Zadatch" which was Cyrillic for "Theory of the Solution of Inventive Problems") [7].

"Empathy" was a technique to become one with the subject being studied such as imagining one-self being inside various processors, actuators and sensors so as to be able to "think" as a bot.

The seven letters from "SCAMPER" stood for changes that ones could apply to an existing product to create a new one:

- S for Substitute – materials, procedures, peoples.
- C for Combine – with other assemblies or processes.
- A for Adapt – change product function, use parts from another product.
- M for Modify – attributes, materials, processes.
- P for Put – putting existing product to other uses.
- E for Eliminate – remove components, simplify systems or processes.
- R for Reverse – reverse functionality or goals.

This technique was applied in class with students to solve a design problem about driving concrete piles into the permanently frozen ground of Siberia as described in [7]. The design challenge was how to resolve a contradiction in requirements: in order to drive the pile satisfactory into the permafrost, it needed to have a pointed tip, however in order to have maximum load bearing capacity it needed a blunt tip instead.

Genrich Altshuller, a Russian engineer working in a patent office, observed that many inventive technical problems from various fields of engineering were solved using the same generic approaches. This realization led him in 1946 to introduce the Teoriya Resheniya Izobretatelskikh Zadatch (TRIZ) offering a systematic approach towards product and process innovation and development [7]. TRIZ is rich in many ideas and techniques that are beyond the scope of this course, therefore only Substance – Field relationship, and Contradiction aspects were presented. First, we discussed the Substance-Field (Sufield) aspects which evolved around energy, tool, and object as shown in Figure 6. This type of relationship was widely found in Technological Systems which existed to perform functions. A function was an intended direct action of a tool on the object. A tool was a component that directly controlled the behavior of an object. This interaction evolved into a system which could be biological, technological, societal, etc… Fey and Rivin pointed out that Technological Systems were organized as hierarchies [14]. In a hierarchy, any system contained subordinate systems and itself might serve as a component for a higher-level system.

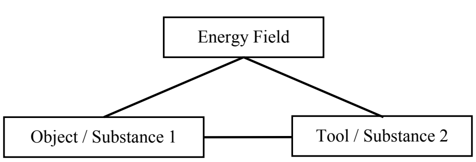


Figure 6. The Triad Relationship.

TRIZ used a Sufield modeling approach for analysis and synthesis of physical structures and processes in technological systems. This modeling technique divided a large problem into smaller pieces to simplify the work into subsystems and their interactions. The term substance could be a material object ranging from a simple pencil to a complex aircraft system. The term field generally referred to fields in physics, such as, electrical, mechanical, optical, or gravitational.

For example, when a robot was designed to kick a ball into a goal, it could be simplified into subsystems. One subsystem could be to find the ball (see or sense) and then aligned the ball with the goal before it kicked it towards the goal with sufficient force. We realized that finding the ball would require sensors which existed in many types. The sensor or sensors that were selected must be able to detect the object and also to discriminate it from any other matter around it. For the sensor to be able to perform its mission correctly, it must be connected to a reliable and proper energy source that enabled its functions. Next would be a mechanical function to move the robot to the ball and align itself to kick the ball into the goal. This subsystem required interactions between three components: a micro-controller (Field), sensor (Tool / Substance) and mechanical movement (Object / Substance). The microcontroller supplied the intelligence and energy required, while the sensor identified the ball and informed the mechanical movement via the microcontroller to kick the ball. This was a subsystem that could be isolated and addressed as needed.

Secondly, we discussed about the TRIZ contradiction matrix. TRIZ heuristics involved criteria, methods, or principles for deciding which among several alternative courses of action promised to be the most effective in order to achieve some goal [15]. The contradiction matrix offered an evaluative method for more than 1200 types of Technical Contradictions formulated from 40 Inventive Principles. A general form of the contradiction matrix included identifying the following:

State design attribute 1
1a. trade-off design attribute
1.a.1 Design Principle (1) to consider
- idea or ideas emerging
1.a.2 Design Principle (2) to consider
-idea emerging (if no idea….say so)
1b. 2nd trade-off design

For example, we used the TRIZ Contradiction Matrix to Innovate Valves in the Valve Train system of an Internal Combustion Engine as follows:

Identified design attributes for valves are:
Speed (9)
Shape (12)
Strength (14)
Reliability (27)
Complexity of control (37)
----

1. Speed (9)
1a. Force (10)
- The other way round (13)

Provide electromagnetic solenoid to actuate and de-actuate the valve. Also an electromagnet would be placed at the other end (position valve assumes at de-actuation state [closed]) to provide added holding power. This innovation would change the current vertical up and down movement into a horizontal configuration. The current valve with a stem will be replaced with one without a stem, etc...

For lack of space, we did not report herein the complete Contradiction Matrix for this valve redesign exercise, but as inferred from the example above, a designer weighed actions and consequences that resulted from the choices made. In this way, a deliberate systematic decision was made for the new innovation.

III. Preliminary Course Assessments

At the time of writing of this manuscript, we are about half way into the Fall 08 semester, thus we can only report on the earlier results from our assessment process. The class started out with 10 students, but now has 8 students.

A. *Student Learning Styles*

Results from our learning styles survey showed that students were mostly visual and kinesthetic learners, with only 2 students predominantly auditory (see Table I).

TABLE I
Visual, Auditory and Kinesthetic Learning Style Scores for students in ENGR-1120-F08

Student	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Visual	27	**37**	**41**	**37**	23	**35**	31	**33**
Auditory	21	31	33	25	29	19	**33**	**33**
Kinesthetic	**41**	**39**	21	29	**35**	23	**33**	23

Although the data showed that 4 students might be kinesthetic learners, during class only 1-2 students were actively taking notes via the Pen Tablets or on paper, while the majority of students would just watch the “show”.

B. *Creativity Surveys*

8 students took the creativity surveys from Gillis and Raudsepp and the results are listed in Table II. Please note that there was not a 1 to 1 correspondence between Table I and Table II regarding any individual student, but that there was a 1 to 1 correspondence between Gillis score and Raudsepp score for any given student in Table II.

TABLE II
Student Scores from Gillis and Raudsepp creativity surveys (ENGR-1120-F08)

Student	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Gillis	6	8	5	7	8	8	8	6
Raudsepp	39	20	13	16	42	21	-13	5

In Gillis’ survey, the scores were to be interpreted as follows:

- 7 to 10 – Creative thinker.
- 3 to 6 – Average creative ability.
- 0 to 2 – Need to follow a pre-set plan.

In Raudsepp’s survey, the scores were to be interpreted as follows:

- 80 to100 – Very Creative.
- 60 to 79 – Above Average.
- 40 to 59 – Average.
- 20 to 39 – Below Average.
- -100 to 19 – Noncreative.

C. *Mid-Term Course Survey*

On October 8th 2008, the current 8 students were given a mid-term survey and 7 students responded.

Students were asked to respond to the following 7 questions using a 6-point Likert scale where "StD" meant "Strongly Disagree", "D" meant "Disagree", "SlD" meant slightly disagree, "SlA" meant "Slightly Agree", "A" meant "Agree" and "StA" meant "Strongly Agree":

1. In-class course materials delivery methods were effective.
2. I understood the materials presented during in-class lectures.
3. In-class materials presented via the second display were effective.
4. Recorded classroom lectures were useful.
5. Pre-recorded narrated tutorials were useful.
6. I felt comfortable going through multi-media presentations on WebCT.
7. I understood the materials presented in recorded lectures and narrated tutorials.

Student responses are shown in Table III. Although results from Questions 1 and 2 showed that in-class materials and delivery methods were mostly effective, one student did not see any benefit of using the second display or the utility of recorded classroom lectures. This student also commented that he or she preferred to take note with pen and paper, as this would not require a computer to review notes and never used the provided pen tablet. This student also listened to only 1 recorded classroom lecture but found the narrated tutorials useful to help doing homework.

TABLE III
In-class & Off-class materials effectiveness survey results (ENGR-1120-F08 – 7 responses out of 8 students)

Question #	**“StD”**	**“SlD”**	**“D”**	**“SlA”**	**"A"**	**"StA"**
1				2	3	2
2			1		5	1
3		1		1	3	2
4		1			2	4
5				1	1	5
6					4	3
7				2	5	

The results for Questions 2 and 7 showed that the effectiveness of the in-class and outside-of-class materials were about the same. The student answering a “Disagree” on Question 2 answered an “Agree” on Question 7. The main author had seen this pattern repeated in his other courses. The interesting fact to point out was that the recorded lectures were just straight unedited recordings of the classroom activities, but somehow the students felt that they understood the materials in the recorded versions better, perhaps because it would be the 2nd time that they would see and hear the materials, and perhaps the "timing" of the review was more conducive to their learning processes as they were "ready" to deal with the materials, and also they could pick out only the needed information as many times as needed, all of which not possible to achieve during the actual lectures. The new thing this year was the narrated tutorials specifically designed to guide students in their homework and term projects (Question 5 had 4 “Strongly Agree” responses). Lastly all students felt comfortable in using WebCT to review recorded lectures and narrated tutorials.

Students were also asked to describe how they had used the various instructional technologies provided to them inside and outside the classroom:

1. The dual-display setup seemed to be accepted by the students as shown in these representative comments:
   “I use the dual displays most of the time because I have trouble switching focus from close to far; the show on screen helps a lot”.
   “Dual displays and projection screen help me see the material better and help me understand faster. I use them every week”.
   “Dual displays are nice since the lecture is right next to me. Pen tablets are hard to use so I don’t really use them”.
   “The dual display is very useful. I do not look up at the projection screen that often. I like to use the pen tablet to take notes, and sometimes I use screen captures. I also use the pre-made notes to write on”.
2. Recorded lectures and narrated tutorials were also found useful by the students as shown below:
   “I use the recorded lectures and narrated tutorials weekly to help with homework. These also make it easier to learn the presented material over the week instead of cramming it into 50 minutes”.
   “Online lectures and the tutorials are good for labs or for preparing for quizzes”.
   “I use recorded lecture to refresh myself on the day of my classes. Narrated tools help a lot for my assignments”.
   “The narrated tutorials and recorded lectures are very helpful. It is good to be able to go back and listen to the lecture in case you miss something in your notes”.

IV. Conclusions

In this report, we had described the curriculum for a robotics-based immigration course into the Computer Systems Engineering BS degree and also a specific mix of instructional technologies to facilitate in-class and outside-of-class learning. When asked this question “What have you liked about the course this semester?” selected responses were:

“Learning about something completely new and different. Robotics is something I never thought I would have experience with”.
“The interactive nature and integrated technology”.
“Very interactive class time with helpful online videos”.
“I have liked learning something interesting but dislikes the difficulty of it”.
“I like learning how to write programs that make real objects do things! I look forward to the robot projects”.

In short, we could say that so far our goals for the course and for the students were met (except for the pen tablets). Some students were without any programming background and thus were quite “stretched” to stay in the course so far. In future offerings of this course, we will ask for a 75 minute weekly time slot instead on the current 55 minute session so that we can have more time for hands-on in class activities for the students and also for instructors to administer regular assessments such as weekly quizzes. As these students are currently freshmen, they will need to be surveyed after graduation about the effectiveness of the skills learned here.

References

[1] S. Yost, M. Krishnan, and M. Paulik, "Development of an Integrated Spiral Curriculum in Electrical and Computer Engineering", *ASEE Technical Paper No. AC 2008-210*, presented at the ASEE Annual Conference in Pittsburgh, Pennsylvania, June 22-25, 2008.

[2] A. O'Fallon, J. R. Hagemeister, C. Cole, and J. Harris, "Increasing Student Retention and Comprehension by Providing Every Student Their Own Industry Standard Tools in Introductory Classes", *ASEE Technical Paper No. AC 2008-2335*, presented at the ASEE Annual Conference in Pittsburgh, Pennsylvania, June 22-25, 2008.

[3] M. Mataric, J. Fasola, and D. Feil-Seifer, "Robotics as a Tool for Immersive, Hands-On Freshmen Engineering Instruction", *ASEE Technical Paper No. AC 2008-139*, presented at the ASEE Annual Conference in Pittsburgh, Pennsylvania, June 22-25, 2008.

[4] C. N. Thai, “A Dual-Mode/Dual-Workspace Classroom Environment", *ASEE Technical Paper No. AC 2008-257*, presented at the ASEE Annual Conference in Pittsburgh, Pennsylvania, June 22-25, 2008.

[5] R. Root-Bernstein and M. Root-Bernstein, *Sparks of Genius*. New York, NY: Mariner Books, Houghton Mifflin Co., 1999, pp. 182-201.

[6] B. Eberle, *Scamper: Creative Games and Activities for Imagination Development*. Waco, TX: Prufrock Press, 1997.

[7] S. Kaplan, *Inventive Problem Solving - Critical Thinking for Problem Solving Using TRIZ.* Fort Wayne, IN: Aptimise, 2008.

[8] Robotis, *Bioloid User's Guide, V 1.1.* Bucheon City, Republic of Korea: Robotis Co., LTD., 2007.

[9] Robotis, *Bioloid Quickstart (Comprehensive), V 1.1.* Bucheon City, Republic of Korea: Robotis Co., LTD., 2007.

[10] K. Bain, *What the Best College Teachers Do*. Cambridge, MA: Harvard University Press, 2004.

[11] E. Raudsepp, “How Creative are You?”, in *Creativity*, edited by A. D. Timpe, Ed. New York, NY: Facts on File Publications, 1987, pp. 34-38.

[12] J. G. Gillis, “Creativity, Problem Solving and Decision Making”, in *Creativity*, edited by A. D. Timpe, Ed. New York, NY: Facts on File Publications, 1987, pp. 209-213.

[13] R. L. Martin, The *Opposable Mind*. Boston, MA: Harvard Business School Press, 2007.

[14] V. Fey and E. Rivin, *Innovation on Demand: New Product Development Using TRIZ.* Cambridge, UK: Cambridge University Press, 2005.

[15] S. D. Savransky, *Engineering of Creativity: Introduction to TRIZ Methodology of Inventive Problem Solving*. Boca Raton: FL: CRC Press, 2000.

Bridging the Emotional Divide in Instructional Design: A Kansei Perspective

Chwen Jen Chen, Kee Man Chuah and Chee Siong Teh
Faculty of Cognitive Sciences and Human Development
Universiti Malaysia Sarawak
Kota Samarahan, 94300 Sarawak, Malaysia
cjchen@fcs.unimas.my, chuahkeeman@gmail.com and csteh@fcs.unimas.my

Abstract- **Emotion has long been regarded as an obstacle to cognitive processes in learning. Despite various studies that showed strong influences of emotional state in learning, instructional designers tend to overlook its role when designing and developing instructional materials. This is mainly due to the lack of proper method or framework in linking emotions and instructions. Thus, with recent renewed interest on affective issues in instructions, this paper offers a fresh perspective by integrating Kansei Engineering methodology in the instructional design process. Kansei Engineering is a proven methodology for translating human feelings into a design. This paper describes how Kansei Engineering methods can be used to elicit important design elements of instructional materials that would in turn induce positive emotions on the learners and optimize effective learning.**

I. Introduction

For decades, emotion has been regarded as a threat to rational thinking. There has been a constant deep-rooted conviction that emotions are unreliable and untrustworthy and that for sanity to prevail, rationality and intellect must function unfettered by the vagaries of emotion [1]. This belief has formed a major influence in the domain of teaching and learning. Learning theories have largely treated emotion and cognition as occupying separate realms and cognitive processes have been given a primary place in the educational scheme of things at the expense of emotions [2]. Though there were attempts to challenge this view over twenty years ago, efforts by researchers such as Martin and Briggs [3] to combine both cognitive and affective domain in creating a more holistic framework for instructional design were seen as problematic and unpopular. Such division between emotion and cognition occurs due to several reasons. The primary reason is the multitude of definitions on the term "emotion" that more often than not conflict with each other [1]. Moreover, the difficulty of research methodologies such as direct observations of private emotional experience and the experiment setup, which often conceal the true nature of emotional experience, is another a major barrier to overcome [4].

Nevertheless, recent development in instructional design and learning sciences has seen a mounting awareness and revitalized efforts among educators and instructional designers on the need to reconsider the role of emotions in instructions. Though limited, several emerging approaches such as FEASP- [4] and ECOLE-approach [5] have been introduced in order to facilitate instructional designers in designing emotionally sound instructional materials (especially computer-based materials) and instructions. A more prominent attempt is done by Astleitner in [3] who have introduced the framework for Emotional Design of Instruction (EDI), which combines the fields of emotional and instructional design. Despite these attempts, emotions have not featured significantly in instructional research and are often overlooked by instructional designers when developing instructional materials mainly due to the lack of proper method in linking emotions and instructions.

Thus, this paper aims to address this gap of emotional divide in instructional design by highlighting the role of emotions in learning and instruction, as well as the need to consider affective and emotional factors when designing instructional materials. In relation to that, this paper proposes the use of Kansei Engineering methodology as an additional tool to facilitate instructional designers in making judgment pertaining to the emotional effects of various design elements.

II. The Role of Emotion In Learning

Generally, emotions include one's action tendencies, desires, feelings, and physiological responses. Kleinginna and Kleinginna [6], in a more comprehensive manner, defined emotion as a complex set of interactions among subjective and objective factors, which can give rise to affective experiences, generate cognitive processes, activate widespread physiological adjustments, and lead to behavior that is usually goal-directed, and adaptive. In learning, emotion may either disrupt or promote information processing. According to Pekrun in [7], emotions have an effect on learning and achievement as mediated by attention, self-regulation and motivation. They direct a person toward or away from learning matters in learning situations, which eventually leads to self-regulated learning.

In addition, several studies have provided empirical evidences that positive emotions have a crucial effect on diverse cognitive processes such as information processing and problem solving. For example, Isen and Reeve in [8] based on their empirical study mentioned that learners who are feeling happy or in positive state of emotions are more cognitively flexible and more able to see potential relations among stimuli than other learners in a neutral state. In another study, Fredickson [9] identified four positive emotions (joy, interest, contentment, and love) that broaden one's scope of attention, cognition and action. She further suggested that the broadening effect triggered by positive emotions builds a range of personal resources that would enhance one's cognitive process. There-

Acknowledgement: The author acknowledges the financial support rendered by UNIMAS through Fundamental Research Grant Scheme FRGS/05(02)/616/2006(49), Ministry of Higher Education, Malaysia.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_33, © Springer Science+Business Media B.V. 2010

fore, it is safe to conclude that positive emotions can indeed promote effective learning. Despite that, the effect of instructional materials and instructions in inducing positive emotions is yet to be explored thoroughly.

III. Emotion And Instructional Design

Traditionally, in the field of instructional design, a strong emphasis is given to learners' cognitive and some motivational processes. As pointed out by Reigeluth [10], emotions have not been sufficiently attended in instruction. One possible factor is that emotions have been conceived as interfering with the cognitive learning objectives and achievement especially when Gagné's nine condition of learning [11] were the main source for instructional designer. In his model, there were no indictors of emotions and the listed instructional events were mainly cognitive in nature.

However, with the propagation of instructional models which are derived from the constructivist approach, instructional designers begin to realise the fact that cognitive, social and emotional development cannot be viewed in isolations as each is closely linked with the other. Wilson [12] further mentioned that instructional designers should not be viewed only as designers of materials. Instead, they should be considered predominantly as designers of experience, specifically learning experience. According to him, on both levels, they move beyond purely technical issues of theory application and enter into the realm of aesthetics that inevitably involves human feelings and emotions.

Although addressing affective and emotional issues is nothing new in the field of instructional design, it is only recently that theoretically guided approaches focusing on instructional strategies to influence learners' emotions are introduced. The most notable one is the FEASP approach as proposed by Astleitner [3]. In the approach, Astleitner suggested that there are five emotions that educators need to consider in the learning context, which are fear, envy, anger, sympathy and pleasure (FEASP). He also added that instructional designers have to consider these emotions in order to optimize the learner's emotion states during the learning process. Astleitner and Leutner in [13] further conceptualized the FEASP approach in guiding the design of computer-assisted instructional materials. They outlined several strategies in identifying emotional problems faced by learners during the use of instructional technology and suggested ways to improve those problems. Another more recent approach is the ECOLE approach (*Emotional and Cognitive Aspects of Learning*) that is rather similar to FEASP approach in which it aims at improving the quality of instruction by increasing positive emotions and achievement and by avoiding negative emotions. However, these approaches are mainly for designing instructions in the classroom.

In bridging the emotional divide in instructional materials, two recent studies [14], [15] have investigated how aesthetic elements such as colours, layout and graphic illustrations of multimedia learning materials could increase the positive emotions of the learners whose emotional state in the beginning of the learning was neutral. The studies showed how students managed to improved their problem solving and decision making ability (as shown in their transfer test results) when given a stimulus that could stimulate cheerful and relaxing feelings. This suggests that positive emotions can be induced from the instructional materials through the quality of the aesthetic design. Both studies have also highlighted the need for instructional designers to incorporate principles of emotional design in instructional design as it is able to affect learners' learning experiences and performance. On close examination of the aforementioned studies, it is questionable how emotions could be categorised to facilitate instructional designer in using the information during the actual design of instructional materials due to the limitation of the employed methodology. As such, due to the systematic procedures in which it is possible to evaluate quantitative relationships between emotions and design elements, Kansei Engineering is seen as a potential method in offering plausible solutions to this problem.

IV. Kansei Engineering

Kansei Engineering or affective engineering is a methodology that assimilates human Kansei (psychological feelings and emotions) into design elements, with an aim to create products or designs that users or consumers will satisfy [16]. This method is derived from the understanding of human Kansei. Kansei is a Japanese term that refers to human's sensitivity, sensibility and feelings, in which when presented with a stimulus, it will be evoked and hence influences the judgement of a person on that stimulus, either positive or negative [17]. The concept of Kansei is then combined with the field of engineering in assisting the development of new products that would closely resemble consumer's insights and desire. Kansei Engineering is based on subjective estimations of products and concept properties and helps users to express their demands on the product including those which they might not be aware of [18].

The process of performing Kansei Engineering methodology can be illustrated in the following diagram (Fig.1):

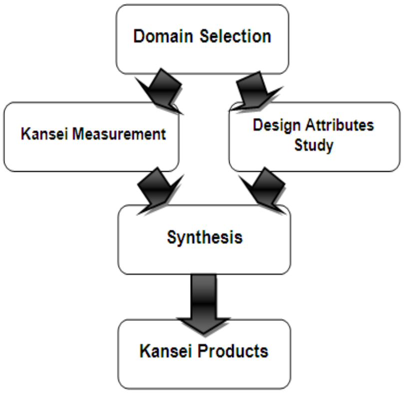


Fig. 1. Principle of Kansei Engineering (Adapted from Nagamachi [18])

A. *Choosing a Domain*

In Kansei Engineering, choosing the domain includes the selection of a target group, market-niche, and specification of the new product. Based on this information, product samples are collected, representing the domain. The Kansei Domain can be understood as the ideal concept behind a certain product. Thus, the major task in this step is to define the domain and find representatives (such as design drawings, products or samples) covering wide part of the domain.

B. *Kansei Measurement*

Kansei measurement is the process of capturing consumer's emotions or internal sensation. Kansei measurement covers physiological measures and psychological measures. Physiological measures target to capture users' behaviors and body responses. On the other hand, psychological measure (which is the popular one between the two) deals with human emotional and mental states. In this psychological measure, users' changes in emotions are captured using self-reporting system such as Semantic Differential (SD) scale or free labeling system. In a typical Kansei measurement procedure, users will be required to rate a product on the Semantic Differential scale (see Fig. 2), which contains list of words in a pre-determined scale range. These words (known as Kansei words) are compiled from various sources such as target users, experts, pertinent literature and the like.

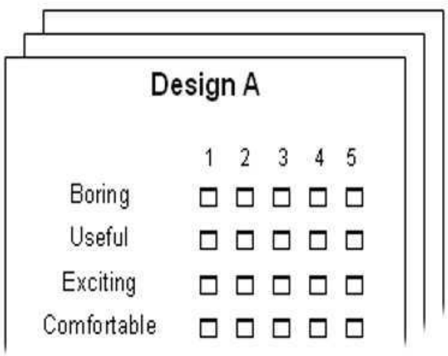


Fig. 2. Example of Semantic Differential scale questionnaire

C. *Design Attributes Study*

In this step, identification of specific design features such as size, shape and colors are conducted. This is to determine the sets of design parameters to be studied and evaluated by the target users. Generally, product or design attributes are selected from the existing products available in the market. In some cases, however, the product attributes can be created or designed from scratch by the product designers especially when there are limited designs available within the selected domain.

D. *Synthesis*

Upon obtaining the evaluation data from Kansei measurement, the correlation between the Kansei words and design attributes (e.g. colour, layout, and size) is then measured using statistical and data visualization method. The most frequently used method is Hayashi's Quantification Theory Type I [17]. This method is a variant of the linear multiple regression analysis that can shows the connection between the properties and each Kansei factor.

V. Kansei Engineering Approach In Instructional Design

Many studies conducted have reported positive results from the use of Kansei Engineering methods in enhancing user's satisfaction on a particular product [17]. Advantages of using KE are that abstract feelings are visualized and made comprehensible. Thus, it may provide a structured support for integrating affective values into product design, especially in early and later stages of the product development process. Despite that, KE methods are not restricted only to the field of industrial and product design. There have been various attempts to incorporate KE in other fields such as e-commerce website design [19] and community design. Its application in instructional design is still yet to be thoroughly examined.

In applying Kansei Engineering methodology in instructional design, the steps taken are illustrated in Fig. 3 and its major steps are explained in the following sections.

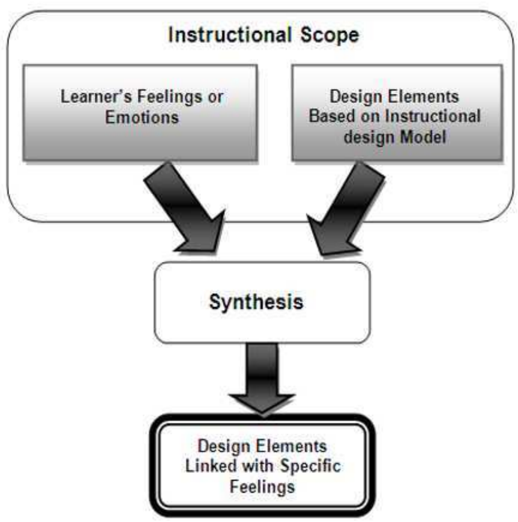


Fig. 3. Proposed approach for Kansei Engineering implementation in instructional design (Adapted from Chuah, Chen and Teh [20])

In designing instructional materials, a designer would commonly select a guiding model based on the scope of the instructional materials. This scope covers the descriptions of target learners, subject matter as well as learning objectives. After an instructional design model is chosen, the designer would proceed by generating design elements based on the prescribe methods or components of the model. To investigate whether the design elements are emotionally sound, it is then presented to the learners for Kansei evaluation. The synthesis process using statistical methods such as Quantification Theory Type I [18] can create links between learner's emotions and design elements and thus form groups of design element linked with specific feelings. In other words, instructional designers will be able to identify the design elements related to the component of the chosen instructional design model that can generate positive emotions. Hence, the final version of the instructional material that suit learners' desire and emotional needs can be created. However, unlike in conventional Kansei Engineering methodology, two important aspects needed close attention when applying the methods to the design of instructional materials: 1) selection of Kansei words and 2) determination of design attributes.

A. Selection of Kansei Words

In Kansei Engineering, user evaluation of a particular design element can be done using various techniques but the most common one is via Kansei survey, which includes a set of Kansei words listed in the form of good-bad manner. Instructional designers can obtain these words from various sources such as pertinent literature, target users, as well as experts. However, unlike in product design, the scope of Kansei words to be used in evaluating instructional learning materials is somewhat different. The selection of Kansei words for such purpose should cover emotions or adjectives closely related to the learning process.

The affective model by Kort, Reilly, and Picard [21] is one example that can guide the Kansei words selection process. The model divides emotions into positive affect and negative affect according to the cognitive dynamics of the learning process. It assumes that there is a process of specific emotions and equivalent cognition in learning context and those emotions are derived from 30 different emotion states. The emotions listed in the model are the common emotions involved in the learning process and it can serve as a guide in generating more Kansei words within the same scope. In addition to that, instructional designers should also be aware that the words selected for Kansei evaluation should be comprehensible to the users in order to obtain feedback that is more accurate and reliable.

B. Determination of Design Attributes

As mentioned in the Kansei Engineering methodology, the design elements are normally selected from the available products in the market. In the case of instructional design, such selection method may not be possible. Instead, in determining the design elements, careful reference to the components within the chosen instructional design model is crucial. In the initial stage, a complete instructional material fulfilling all the components in a chosen instructional model is developed. Then, depending on the intent of the instructional designer, design elements related to each component will be "manipulated" or "removed" accordingly. This will thus create a set of designs with different highlighted component for user evaluation. Target learners are required to rate all these design sets using the Semantic Differential questionnaire given to them. Changes to the learners' emotional state can then be identified by analysing the data collected using multivariate analysis. The output from the analysis is sets of design elements linked to specific feelings or emotions. This will provide valuable information for the instructional designers to determine which design elements that can be used to induce positive emotions.

VI. Conclusions

It can be noted from the review of literature (such as in [8] and [9]) that emotions played a major role in promoting effective and engaging learning. Nonetheless, studies that investigated emotions in instructional design were still insufficient and did not address the emotional divide in instructional materials. As such, this paper proposed the use of Kansei Engineering methodology in instructional design particularly in the design of instructional materials. It attempts to provide an alternative perspective towards inducing emotions in instructions. It is reckoned that Kansei Engineering methodology has the potentials to reduce the gap that exists between cognition and emotion in the process of designing and developing instructional materials regardless of its medium or technology. In addition, the possible connection between KE methodology and instructional design is useful in helping instructional designers to create emotionally sound materials that facilitate learning.

References

[1] L. O'Regan, "Emotion and e-learning," *Journal of Asynchronous Learning Networks*, vol. 7, no. 3, pp.78-92, 2003.

[2] M. Martinez, "Key design considerations for personalized learning on the Web", *Educational Technology and Society*, vol. 4, no.1, pp.26–40, 2001.

[3] B. Martin and L. Briggs, *The Affective and Cognitive Domains: Integration for Instruction and Research,* Englewood Cliffs: Educational Technology Publications, 1986.

[4] H. Astleitner, "Designing emotionally sound instruction: The FEASP-approach," *Instructional Science*, Vol. 28, pp.169-198, 2000.

[5] M. Glaser-Zikuda, Fub, S., Laukenmann, M, K. Metz, and C. Randler. "Promoting students' emotions and achievement—Instructional design and evaluation of the ECOLE-approach", *Learning and Instruction*, vol. 15, pp.481-495, 2005.

[6] P. Kleinginna and A. Kleinginna, "Categorized list of emotion definitions, with suggestions for a consensual definition," *Motivation and Emotion*, vol. 5, pp.345-379, 1981.

[7] R. Pekrun, "The impact of emotions on learning and achievement: towards a theory of cognitive/emotional mediators," *Applied Psychology: An International Review*, vol. 41, pp.359-376, 1992.

[8] A.M. Isen and J. Reeve, "The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behaviour, and self-control," *Motivation and Emotion*, vol. 29, no. 4, pp.297-325, 2005.

[9] B.L. Fredickson, "What good are positive emotions?," *Review of General Psychology*, vol. 2, no. 3, pp.300-319, 1998.

[10] C.M. Reigeluth, "Instructional theory, practitioner needs, and new directions: some reflections," *Educational Technology*, vol. 37, pp.42–47, 1997.

[11] R.M. Gagné, *The Conditions of Learning (2nd ed.),* New York: Holt, Rinehart and Winston, 1970.

[12] B.G. Wilson, "Broadening our foundation for instructional design: Four pillars of practice," *Educational Technology*, vol. 45, no. 2, pp. 10-15, 2005.

[13] H. Astleitner and D. Leutner, "Designing instructional technology from an emotional perspective," *Journal of Research on Computing in Education*, vol. 32, pp. 497-510, 2000.

[14] S. Park and J. Lim, "Promoting positive emotion in multimedia learning using visual illustrations," *Journal of Educational Multimedia and Hypermedia*, vol. 16, no. 2, pp. 141-162, 2007.

[15] E. Um, H. Song and J. Plass, "The effect of positive emotions on multimedia learning," in *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007*, C. Montgomerie and J. Seale, Eds. VA: AACE, 2007, pp. 4176-4185

[16] M. Nagamachi, "Kansei Engineering: The implication and applications to product development," *IEEE International Conference*, 1999.

[17] S. Schütte, *Engineering emotional values in product design - Kansei Engineering in development*, Linköping Universitet: Sweden, 2005.

[18] M. Nagamachi, "Kansei Engineering: A new ergonomic consumer-oriented technology for product development," *International Journal of Industrial Ergonomics*, vol. 15, pp.3-11, 1995.

[19] M.L. Anitawati and M.N. Nor Laila, "Kansei Engineering Concept in E-commerce Website", *International Conference on Kansei Engineering and Intelligent Systems 2006*, Aizu-Wakamatsu, Japan, 2006, pp. 117-124

[20] K. M. Chuah, C. J. Chen and C. S. Teh. "Incorporating Kansei Engineering in instructional design: Designing virtual reality based learning environments from a novel perspective," *Themes in Science and Technology Education*, vol. 1, no. 1, pp. 37-48, 2008

[21] B. Kort, R. Reilly, and R.W. Picard, "An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion," *Proceedings of ICALT*, 2001.

Fuzzy Knowledge Processing for Unveiling Correlations between Preliminary Knowledge and the Outcome of Learning New Knowledge

Sylvia Encheva and Sharil Tumin

Abstract— **Socratic dialogs apply a dialectic method to present questions to a learner and let the learner reconsider his or her own thinking. Further more, Socratic dialogs foster critical thinking on course content through a series of questions and answers that enable learners to explore and develop understanding.**
All this makes Socratic dialogs very useful for automated discovering of individuals who will not be able to obtain sufficient knowledge in new subjects because of a pour theoretical and practical background.

***Index Terms*—Intelligent infrastructures and automated methods, logic**

I. Introduction

A Socratic dialog applies a dialectic method to present questions to the student and lets the student reconsider his or her own thinking. Results show that the Socratic-dialog approach is of great help to students.

The work presented in [11] elaborates the point that informal, quasi-empirical mathematics does not grow through a monotonous increase of the number of indubitally established theorems but through the incessant improvements of guesses by speculation and criticism, by the logics of proofs and refutations.

Refutation by counterexamples depends on the meaning of the terms in question. We may achieve an agreement on the meaning of our terms by defining the term where communication broke down. One of the reasons for that may be lack of familiarity with a particular concept. Familiarity with a concept is defined in [11] as the ability to distinguish a thing which belongs to a predefined class of objects from a thing which does not – known as the extension of the concept of that thing.

Manuscript received October 5, 2008.
S. Encheva is with Faculty of technology, Business and Maritime Science,Stord/Haugesund University College, Bjørnsonsg. 45, 5528, Haugesund, Norway (phone: +47 52 70 26 85; fax: + 47 52 7026 01; e-mail: sbe@ hsh.no).
S. Tumin is with the IT Department, University of Bergen, PO Box 5020 Bergen, Norway (e-mail: edpst@it.uib.no).

Fuzzy reasoning methods are often applied in intelligent systems, decision making and fuzzy control. Some of them present a reasoning result as a real number [9] and [10], while others use fuzzy sets, [8]. Fuzzy reasoning methods involving various fuzzy implications and compositions are discussed by many authors, f. ex. [1], [2], [3], [5], [6], [11], and [13] to name but a few. An interesting prediction method presented in [7] applies formal concept analysis and fuzzy inference.

This work is devoted to development of an automated way for discovering of individuals who will not be able to obtain sufficient knowledge in new subjects because of a pour theoretical and practical background.

The rest of the paper is organized as follows. Section II contains definitions of terms used later on. Section III describes how we propose to select students that need extra classes and Section IV is devoted to a system description. Section V contains the conclusion of this work.

II. Background

Socratic dialogs foster critical thinking on course content through a series of questions and answers that enable students to explore and develop understanding.

A mutually respectful environment and probing questions are essential elements of an effective dialog. Use of multiple question types increases effectiveness. Relationships between question types and their usefulness are summarized in [14] as

Asking about	*Stimulates*
Purpose	Task clarification
Information	Sources and quality investigation
Asking about	*Stimulates*

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_34, © Springer Science+Business Media B.V. 2010

Assumption	Carefulness in circumstances evaluation
Implication	Thoughts control
Point of view	Development of broad mind
Relevance	Elimination of non-critical data
Accuracy	Truth values investigation
Precision	Learning the importance of details
Consistency	Learn to discover contradictions
Refocusing	Turning attention to a related issue (a summation of the dialog's progress before refocusing can increase the clarity of the conversation)

Let P be a non-empty ordered set. If $sup\{x,y\}$ and $inf\{x,y\}$ exist for all $x,y \in P$, then P is called a *lattice*, [3]. In a lattice illustrating partial ordering of knowledge values, the logical conjunction is identified with the meet operation and the logical disjunction with the join operation.

A *concept* is considered by its *extent* and its *intent*: the *extent* consists of all objects belonging to the concept while the *intent* is the collection of all attributes shared by the objects [3] and [12].

A *context* is a triple (G, M, I) where G and M are sets and $I \subset G \times M$. The elements of G and M are called *objects* and *attributes* respectively.

For $A \subseteq G$ and $B \subseteq M$, define

$$A' = \{m \in M \mid (\forall g \in A)\ gIm\},$$

$$B' = \{g \in G \mid (\forall m \in B)\ gIm\},$$

so A' is the set of attributes common to all the objects in A and B' is the set of objects possessing the attributes in B. Then a *concept* of the context (G, M, I) is defined to be a pair (A, B) where $A \subseteq G$ and $B \subseteq M$, $A' = B$ and $A' = B$. The *extent* of the concept (A, B) is A while its *intent* is B. A subset A of G is the extent of some concept if and only if $A'' = A$ in which case the unique concept of which A is an extent is (A, A'). The corresponding statement applies to those subsets B of M which are the intent of some concept.

The set of all concepts of the context (G, M, I) is denoted by $B(G, M, I)$. $\langle B(G, M, I); \leq \rangle$ is a complete lattice and it is known as the *concept lattice* of the context (G, M, I).

For concepts

$(A1, B1)$ and $(A2, B2)$ in $B(G, M, I)$ we write

$$(A1, B1) \leq (A2, B2),$$

and say that

(A1, B1) is a subconcept of (A2, B2),

or that

(A2, B2) is a superconcept of (A1, B1),

if

$A\_1 \subseteq A\_2$ which is equivalent to $B\_1 \subseteq B\_2$

The following two statements from [4] are widely used with respect to concept lattices

Lemma

Assume that *(G, M, I)* is a concept and let

$$A, Aj \subseteq G \text{ and } B, Bj \subseteq M, \text{ for } j \in J.$$

Then

(i) $A \subseteq A``$,
(ii) $A1 \subseteq A2 \Rightarrow A1` \supseteq A2'$
(iii) $A` = A```$
(iv) $(\cup_{j\in J} Aj)` = \cap_{j\in J} Aj`$
(i)` $B \subseteq B``$,
(ii)` $B1 \subseteq B2 \Rightarrow B1` \supseteq B2`$
(iii)` $B` = B```$
(iv)` $(\cup_{j\in J} Bj)` = \cap_{j\in J} Bj`$

The fundamental theorem on concept lattices [4] states

Teorem

Let *(G,M, I)* be a context. Then $\langle B(G, M, I); \leq \rangle$ is a complete lattice in which join and meet are given by

$$\vee_{\in J}\ (Aj, Bj) = ((\cup_{j\in J} Aj)``,\ (\cap_{j\in J} B),$$

$$\wedge_{j\in J}\ (Aj, Bj) = (\cap_{j\in J} Aj),\ (\cup_{j\in J} B)``),$$

Conversely, if L is a complete lattice then L is isomorphic to $\langle B(G, M, I); \leq \rangle$ if and only if there are mappings

$$\gamma : G \rightarrow L$$

and

$$\mu : M \rightarrow L$$

such that $\gamma(G)$ is join-dense in L, $\mu(M)$ is meet-dense in L and gIm is equivalent to $\gamma(g) \leq \mu(M)$ for each $g \in G$ and $m \in M$. In particular, L is isomorphic to $\langle B(G, M, I); \leq \rangle$ for every complete lattice L.

The *sum-of-1-criterion* [6] states that

$$\Sigma_{i \in M_i} \mu_i(x) = 1, \ \forall \ x \in \chi$$

where M_i, $i = 1, ..., k$ denotes all possible membership terms $\{m_i, i = 1, ..., k\}$ of a fuzzy variable in some universe of discourse χ.

An affiliation value α to a concept (A, B) is defined as

$$\alpha(\mathrm{A}, \mathrm{B}) = \frac{\sum_{o \in \mathrm{A}, f \in \mathrm{B}} m_{of}}{|\mathrm{A}|.|\mathrm{B}|}$$

The affiliation value represents the relative extent to which an object belongs to this concept or an attribute is common to all objects in the concept.

In the derived graph, i.e. in the concept lattice each vertex represents a concept from the table in Fig. 1. The concepts are arranged hierarchically in this concept lattice, i.e. the closer a concept is to the supremum, the more attributes belong to it. Moving from one vertex to a connected vertex which is closer to the supremum means moving from a more general to a more specific description of the attributes if an object appears in both concepts. The number of attributes increases and the number of attributes decreases along such an edge.

By θ we denote the threshold for membership values above which an entry is regarded as significant, as in [6]. This is achieved by computing the arithmetic mean of all entries within a column and take it as a threshold.

III. Dialogs

Socratic dialogs stimulate students' creative thinking, contribute for faster development of problem-solving skills and help refinement of students' decision-making criteria.

Socratic dialogs are further discussed in [9] with respect to the following

Creative thinking:

Questions drive the creative process. Socratic dialogs are an excellent way to challenge your students to think deeply and creatively about the course content. In addition, this type of discourse prepares students for dialogs in the workplace.

Problem Solving:

With Socratic dialog an instructor can effectively model problem-solving skills using questioning to walk students through the problem-solving process. Answering questions of purpose and relevance will help students understand and define a problem. Answers to information and interpretation questions will generate possible solutions. These solutions can be further examined by answering questions of precision and consistency.

In addition, Socratic dialog can help students overcome obstacles they have with problem solving. Answering point-of-view questions will help students realize the validity of multiple perspectives. Faulty assumptions can be uncovered by answering questions of assumption.

Decision Making:

Socratic dialog can be used to help students develop decision-making criteria. Questions of information, interpretation, and implication can help students identify and weigh the various factors that comprise their decision-making criteria. Answers to point-of-view, consistency, and precision questions can help students refine the criteria.

Thus Socratic dialogs are proposed to be used for discovering students who will not be able to obtain sufficient knowledge in a new subject because lack of some basic knowledge.

Earlier experience in teaching concepts is used as reasoning set for establishing whether a new student would have serious difficulties in the current learning process. The summarized experience is first reworked as to be placed in a table relating objects and attributes. This work is followed by building a concept lattice visualized as a Hasse diagram. Fuzzy functions are then applied for deciding on the final recommendation.

By $F(s)$ we denote failure of a particular student to learn a concept. The attribute $F(s)$ has values calculated according to the function

$$F(s) = \begin{cases} \sum F(C_i).\alpha_{Ci}, \sum F(C_i).\alpha_{Ci} < 1 \\ 1, otherwise \end{cases}$$

where Ci are the concepts that have the same F values as the student and α_{Ci} are the corresponding affiliation values. If the student's results are not equal to any of the known attributes' we take the concept that has the closet attribute value.

IV. System implementation

A system prototype can be build as a Web-based application using *Apache* HTTP server, *mod_python* module and *SQLite* database. The *mod_python* module provides programmable runtime support to the HTTP server using Python programming language. The whole application components;

- Web-based users interface,
- application logic, and
- database interaction written in Python.

Python provides a programming environment for implementing script-based handler for dynamic content, data integration and users' software agents.

The back end SQLite databases are used to store both static and dynamic data.

Apache is a modular Web server that can incorporate a high level scripting language as a module such as f. ex. mod_python.

Using mod_python, python interpreter becomes a part of the Web server. SQLite is a small footprint, zero-administration and serverless database system. SQLite stores persistence data into files. SQLite thus provides a database platform for multiple databases.

V. Conclusion

The suggested approach can be applied to various subjects that require preliminary knowledge and or skills.

References

[1] E. Bellman and L.A.Zadeh, *Decision making in a fuzzy environment*, Management Sciences, Ser. B, 17, 1970, pp. 141-164

[2] C. Carlsson and R. Fuller, *Optimization under the fuzzy if-then rules,* Fuzzy Sets and Systems, 1998

[3] Chang, Kuo-En; Lin, Mei-Ling; Chen, Sei-Wang, *Application of the Socratic Dialogue on Corrective Learning of Subtraction*, Computers and Education, v31 n1 p55-68 Aug 1998

[4] B. A. Davey, and H. A Priestley, *Introduction to lattices and order*, Cambridge University Press, Cambridge, 2005

[5] R.Felix, *Relationships between goals in multiple attribute decision making*, Fuzzy sets and Systems, vol. 67, 1994, pp. 47-52.

[6] R. Fuller and H.-J. Zimmermann, *Fuzzy reasoning for solving fuzzy mathematical programming problems*, Fuzzy Sets and Systems, vol. 60, 1993, pp. 121-133

[7 C. S. Herrmann, *Fuzzy logic as inferencing techniques in hybrid AI-Systems*, Lecture Notes in Computer Science, vol. 1188, 1997, pp.69-80

[8] C. S. Herrmann, S. Holldobler, A. Strohmaier, *Fuzzy conceptual nowledge processing*, 1996 ACM symposium on applied computing, 1996 pp. 628-632

[9] http://fibers.philau.edu/ntc/I99P01/Socratic\%20Dialog.html

[10] D.K Knox, Socrates: *The First Professor, Innovative Higher Education*, vol. 23, no2, p. 115-26,Winter 1998. Describes Factors to consider when developing a series of questions for a Socratic Dialog. also contains a comprehensive listing of question types.

[11] I.Lakatos, *Proofs and refutations, the logic of Mathematical Discovery*, Cambridge University Press, 1976

[12] E. H. Mamdani, *Applications of fuzzy logic to approximate reasoning using linguistic synthesis*, IEEE Transactions on Computers, vol. 26, No. 12, pp. 1182-1191, 1977.

[13] M. Sugeno, *Fuzzy measures and fuzzy integrals: a survey*, (M.M. Gupta, G. N. Saridis, and B.R. Gaines, editors) Fuzzy Automata and Decision Processes, pp. 89-102, North-Holland, NY, 1977.

[14] M. Sugeno, *Industrial applications of fuzzy control*, Elsevier Science Pub. Co., 1985.

[15] T.Takagi and M.Sugeno, *Fuzzy identification of systems and its applications to modeling and control*, IEEE Transactions on Syst. Man Cybernet.,1985, 116-132.

[16] R. Wille, *Concept lattices and conceptual knowledge systems*, Computers Math. Applications, vol. 23(6-9), 1992, pp. 493-515

[17] L.A.Zadeh, *The concept of linguistic variable and its applications to approximate reasoning*, Parts I, II, III, Information Sciences, 8(1975) 199-251; 8(1975) 301-357; 9(1975) 43-80.

[18] http://www.engsc.ac.uk/journal/index.php/ee/article/view/62/97

Accuracy Improvement Evaluation in Sensorless Dc Motor Speed Estimation by Combining the Dynamic Motor Model and the Ripple Component Detection

Ernesto Vázquez Sánchez[1], Jaime Gómez Gil[1], Miguel Rodríguez Álvarez[2]

[1]Department of Signal Theroy, Communications and Telematic Engineering, University of Valladolid, Spain
[2]Lear Corporation Holding Spain, Carrer Fusters, 54, 43800 Valls (Tarragona), Spain

***Abstract*- Automation of some tasks requires the speed control of dc motors, this also requires a speed observer. In this work it is described a method that combines two sensorless speed techniques for speed estimation without sensors, one of them is based on the dynamic motor model and the other one is based in the ripple component. It is also evaluated the improvement introduced by combining both techniques using as reference the accuracy provided by a method based only on the ripple component. The results clearly show, that the method based on both sensorless speed techniques provides important improvements in the detection of the speed, when the speed is small or when it is changing.**

I. Introduction

Motors are widely used in the industry world, as in domestic and civil applications. Inside these fields, where it is increasing their importance is in the automobile industry. In this sector, it is used for the movement of power windows, sunroofs, doors, etc. [2].

In the majority of applications with dc motors, it is necessary a closed loop speed control, which requires a speed observer [3]. Conventional observers, known as angular speed sensors, are: tachometers, encoders, hall-effect sensors, etc. These observers have the disadvantage that they are an independent element that must be attached to the motor shaft, in a place where it was preferable to put the load. As it is an element different of the motor it increases the failures and the system costs [1].

There are also, other speed observers that are not composed by external mechanical elements. These, known as sensorless speed observers, estimate the speed from the voltage and/or current of the dc motor. Inside them there are two groups [3]. The first group is based on the dynamic dc motor model [4]. They mainly use the electromotive force EMF for the detection. We can find this kind of estimators in [4], [5], [6], [7], [8], [9], [10], [11] and [12]. The model used in these works is lineal and dependant on the motor parameters. Those parameters, have the inconvenience that they change with the working conditions of the motor, this introduces some errors in the measurements. Although those parameters can be dynamically estimated as it is extracted from [13], this usually leads to a non lineal model that increases the computational cost.

The second class of speed observers, without mechanical elements, is based on the ripple component of the motor current. This component, that provides information about the motor speed, is known in the literature as ripple current. The current of a dc motor is composed by a continuous component, plus an alternating component or ripple. The most interesting characteristic of the ripple component is the frequency, which is related with the motor speed through (1). Where f is the frequency of the ripple component, p is the number of pairs of poles of the dc motor, k_d is the number of commutator segments, n is the angular speed of the motor in r.p.m. and η is the maximum common divisor of 2p and k_d [14].

$$f = 2pk_d n/(60\eta) \tag{1}$$

In this case, the frequency depends on dc motor constructive parameters, that are defined when the motor is manufactured. With this kind of observers the problem is usually the ripple frequency detection.

In this work it is implemented a method for the detection of the speed based on sensorless speed estimators. In this method, it is combined the detection based on the ripple component and the detection based on the dynamic model, in order to overcome the deficiencies of each one. Techniques based in the dynamic model have the problem of the dependency on the motor parameters, which change a lot when the external conditions change, for example the temperature. On the other hand, methods based on the ripple component are usually focused on the detection of relative maximums of zero-crosses of the signal (when it has been previously eliminated the continuous component), each one of them is related to a commutation, or ripple component period. With that information they estimate the ripple frequency, and with the help of (1) they obtain the dc motor speed. The detection of maximums or zero-crosses has the inconvenience of the noise, due to power supply perturbations or double commutations. This makes that sometimes inexistent commutation events are detected or the real ones are not correctly detected.

In order to eliminate the inexistent commutations and recover the non detected ones, it is performed an estimation of the motor

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_35,

speed using the dynamic model. With this estimation, it is calculated the most probable commutation period. In parallel, using a ripple detector, there are obtained the current commutation instants. With the help of the most probable commutation period it is possible to filter double commutations; if two pulses are not separated a distance equals to the most probable commutation period plus a certain tolerance, they are ignored. Otherwise, if after a pulse there is no other pulse after the distance of the most probable commutation period, including the tolerance, it is supposed that there is a commutation pulse but due to the noise it could not be detected, therefore a pulse is added in the most probable instant. Finally, with the commutation pulses considered as correct it is calculated the real commutation period, and with it, the ripple frequency and the motor speed.

In next sections it is detailed the employed method and its accuracy, comparing it with the accuracy of the method that is based only on the ripple component. Therefore, the objectives are explained in II, the methods used for the implementation of the system are detailed in III, the algorithm of the method that is being evaluated and used for the comparisons is explained in IV, the results of the study are shown in V, and finally in VI the results are analyzed.

II. Objectives

The objective of this study is to analyze the accuracy improvement that is achieved by combining two types of sensorless speed observers for the detection of the speed in dc motors. There are two different sensorless speed techniques, based on the dynamic model of the motor, and based in the ripple component of the motor current.

The improvement will be measured using as reference the method based only in the ripple component. The accuracy will be obtained analyzing the differences between the detected speed and the real speed of the motor. The error will be measured in two different scenarios, with constant speed, with constant acceleration and with a speed step. In order to get some independency from the employed motor, the tests will be performed with several dc motors.

III. System Description

The hardware schematic of the system is shown in Fig. 1. In this schematic it is shown the dc motor, in this case two different motors are used whose characteristics are shown in Table I. The next element that is shown is the current sensor, it is a 20 mΩ current shunt. Other element is the low cost data acquisition card NI USB-6008, with a maximum sample speed of 10kHz. The card has 4 analog inputs configured in differential mode, these inputs can be configured to have a range from ±1 a ±20 V. One of these inputs is used to sample the voltage of the motor and other input is used to measure the current with the shunt.

The data acquisition card is connected to a PC that is used to process the signals and estimate the speed of the dc motor. The PC is a laptop with a T8300 microprocessor, 3GB of RAM and 320GB of hard disk space. The operating system is Windows Vista, and the development environment is LabVIEW 8.5. In Fig. 2 it is possible to see an image of the explained elements.

In order to detect the real speed of the motor, and make the comparison of the results provided by both methods, it is used a high resolution encoder that is attached to the motor shaft. The encoder is connected to a 32 bit counter in the data acquisition card that is used to count the pulses produced in a fixed time interval, with this value is possible to calculate the speed of the motor. These connections are not shown in the previous schematics.

IV. Methods

The main objective of this work is to evaluate the improvement in the accuracy that is obtained by combining two sensorless speed techniques. In order to estimate the improvement, it will be compared the accuracy obtained with the method that combines both sensorless speed techniques, with the obtained with the method that only uses the technique based on

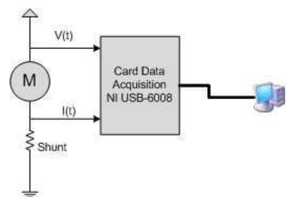


Fig. 1. Schematic of the interconnection of the different elements.

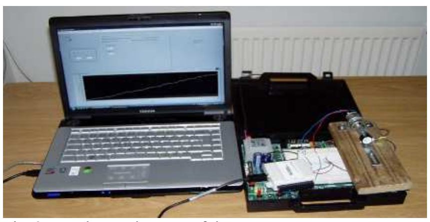


Fig. 2. Hardware elements of the system.

TABLE I
SPECIFICATIONS OF THE DC MOTROS

Parameter / Motor	EMG30	718E385
Nominal Voltage	12 V	12 V
No Load Current Consumption	530 mA	600 mA
Nominal Speed	3000 r.p.m.	5000 r.p.m.
Resistance (R_a)	1.8 Ω	1.5 Ω
Electromotive Force Constant (c)	0.0178 V / r.p.m.	0.00101 V / r.p.m.

the ripple component. To obtain objective results, the block that detects the motor commutations will be the same in both methods.

In the next paragraphs both methods will be described with more details. First it is explained the method based on the ripple component and after that the method that combines both sensorless speed techniques.

A. Method based on the ripple component

The implemented method for speed detection was proposed by Steven Weiss [16]. The method tries in a first stage to detect the commutations produced in the current using the zero-cross technique. In that technique the instantaneous current value is compared with the average value. The comparator produces a positive edge when there is a commutation in the current. Using the time instant of the commutation, the period of the signal and the ripple frequency are calculated and translated to the motor speed.

The block diagram of the method is shown in Fig. 3, it is possible to observe the Ripple Detector and the Speed Detection Unit. The Ripple Detector is used to detect the current commutations. The Speed Detection Unit is used to calculate the period of the commutations and using it, the ripple frequency and the motor speed.

1) Ripple Detector Block

The aim of this block is to detect the commutation instants produced in the current. The schematic of this block is shown in Fig. 4. The current passes through a maximum and minimum peak detector consecutively. The peak detectors each certain time decrease the detected peak value, in order to dynamically adjust the detection to the maximums and minimums of a small interval.

After that it is calculated the average value I_{med} of the maximum and minimum, that is compared with the value of the current. The comparison is made with a hysteresis cycle comparator, in order to eliminate the influence of the small current perturbations. Finally, the signal that goes out the comparator is passed through a positive edge detector that checks if the current sample has a high value and the previous sample a small value. If this condition is met, it means that an edge has been detected and the edge detector puts on its output a *TRUE* value during a clock cycle.

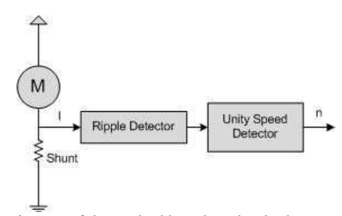


Fig. 3. Block Diagram of the method based on the ripple component.

It must be highlighted that the system works at sample level, as the processing is performed in discrete time. Because of that, when the edge detector produces a *TRUE* value, it means that in the time instant related to the sample that is being processed, a commutation has been produced.

2) Speed Detection Unit

The function of this block is to calculate the speed of the motor using the time instants of the commutations. The flowchart of this block is shown in Fig. 5. The first step is to check if a commutation has been detected; if there is no commutation then there is anything to do. However, if a commutation has been detected, the current commutation instant is added to the commutation instant list T_{k+1}, where k is the number of commutation instants detected.

After that the ripple frequency f is calculated, in order to do that it is possible to use (2). The problem of getting the ripple frequency with this method is that the result has a very nervous and variable behavior.

$$f = 1/(T_{k+1} - T_k) \tag{2}$$

In order to solve this problem, a new estimator is used. It is based on the average calculation during a time T_p with (3), where N is the number of samples used in the average and is defined by (4), f_k is the ripple frequency calculated in the last detected commutation and f_{k+1} is the ripple frequency that is being calculated. The operation [·] is the integer part operator. With this solution it is obtained a speed without perturbations and a good response, following the speed variations, if T_p is correctly adjusted.

$$f_{k+1} = N / \left(\sum_{i=0}^{N-1} T_{k+1-i} - T_{k-i} \right) \tag{3}$$

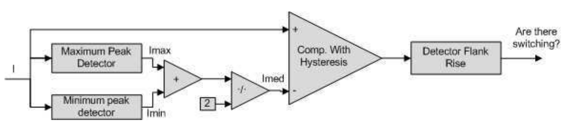


Fig. 4. Ripple detector schematic.

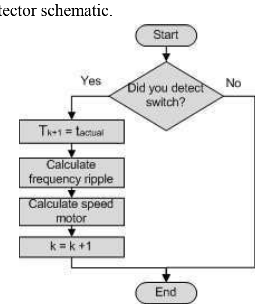


Fig. 5. Flowchart of the Speed Detection Unit.

$$N = [T_p \cdot f_k] \tag{4}$$

Once the ripple frequency is obtained it is possible to calculate the motor equivalent speed using (1). Finally, the number of detected commutations k is updated, adding one additional commutation.

B. Method that combines both sensorless speed techniques

The method used for combining both types of sensorless speed observers is described by Erwin Kessler [15].

The block diagram of the method is shown in Fig. 6. In that block diagram it is possible to distinguish four blocks. Ripple Detector, State Model of the Motor, and Evaluation Unit.

1) Ripple Detector Block

Regarding the implementation of this block Erwin Kessler does not say anything; he only comments that it must be able to detect the commutation instants. The function of this block is the same that for the Ripple Detector block in the method that only uses the technique based on the ripple component, for this reason this block has the same implementation.

2) Motor State Model Block

This block estimates the speed of the motor using the simplified dynamic equation of the motor. That equation, is shown in (5), where V is the voltage across the motor, I is the current, n is the speed, R_a is the resistance, L is the inductance and c is the electromotive force constant. As it is a dc motor, and the value of L is very small, is possible to eliminate the related term. After the simplification of the previous equation and solving for the speed it is obtained the equation (6). This equation is implemented by this block and provides an estimation of the speed.

$$V = R_a \cdot I + L\frac{dI}{dt} + c \cdot n \tag{5}$$

$$n = (V - R_a \cdot I)/c \tag{6}$$

3) Evaluation Unit Block

The evaluation unit combines the information from the two previous blocks. The main function is to process the commutation instants identified by the Ripple Detector. In order to do that, this block removes the non valid commutation instants and adds the non detected commutation instants. Each valid commutation instant is used to calculate the motor speed.

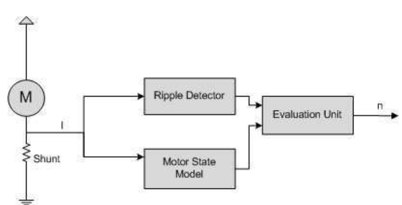


Fig. 6. Block diagram of the method based on the ripple component.

The flowchart is shown in Fig. 7. In order to describe it, we have supposed that the most probable next commutation instant τ_{k+1} and the temporal interval Δt are known. The first step is to check if in the current time instant the Ripple Detector has detected a commutation. If it was detected, the algorithm checks that the current instant is within the interval $\tau_{k+1} \pm \Delta t$. If the previous condition is not accomplished, then it is supposed that the commutation is due to the noise and is ignored. If the condition is accomplished, the commutation is considered valid and is counted, $T_{k+1} = t_{actual}$. With the new information the speed is updated and the new most likely instant is calculated again. If there had not been any commutation, the algorithm checks if the interval where the commutation is expected has been overcome, ¿$t_{actual} > \tau_{k+1} + \Delta t$? If it have been overcome, it means that the Ripple Detector have not detected any commutation in the allowed interval. In this situation, it is supposed that there might be a commutation in that interval, but due to the noise it has not been detected. Therefore it is used as commutation instant the most probable, $T_{k+1} = \tau_{k+1}$. With this value a commutation instant has been detected so it is possible to update the speed and calculate most likely instant for the next commutation.

Once the new valid commutation instant T_{k+1} is known, the speed is updated. In a first step the ripple frequency is calculated using (3) and then the motor speed with (1).

When the speed has been updated, the most likely instant for the next commutation is calculated. For these calculations it is used the information of the ripple frequency previously calculated, and the real instant of the last commutation. All of that is combined using (7), where τ_{k+2} is the most likely instant

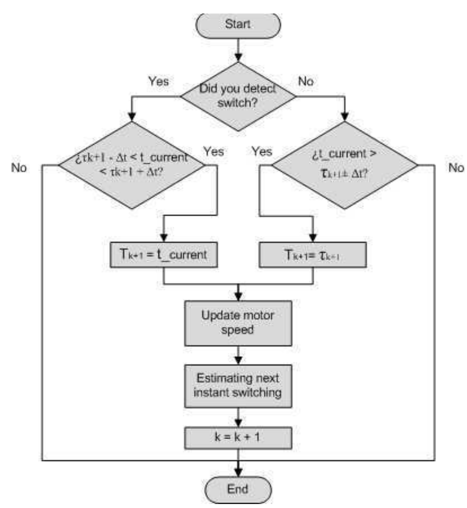


Fig. 7. Flowchart of the Evaluation Unit.

for the next commutation, T_{k+1} is the instant of the last commutation and f_{est} is the ripple frequency related to the speed of the motor n_{est} estimated by the Motor State Model. The relation between them is given by (1). When this point is reached the k index is updated by increasing it one unit.

$$\tau_{k+2} = T_{k+1} + 1/f_{est} \quad (7)$$

The parameter Δt indicates the maximum deviation of the commutation instant with the estimated. That parameter is an indicator of the noise filtering. A high value indicates that many commutations due to the noise have been considered good; a low value indicates that very few commutations due to the noise will be detected. The value of Δt is a parameter that is established before starting the observer. Its value must be under the half of the ripple period related to the maximum working speed of the motor, with this value it is possible to assure that in all the possible situations some noise will be filtered.

V. Results

Once the characteristics of the different methods have been described it is necessary to show the obtained results and the accuracy of each method. Regarding the accuracy, it is shown the improvement that is achieved with the method that combines both sensorless speed techniques. The next paragraphs show the results obtained for each motor separately.

In each test scenario it is shown the average error and the deviation for different constant speeds. The error and the delay when the speed is changing with constant acceleration have been also measured. Finally, a speed step has been produced and it has been measured the time required for reaching the final speed value.

In the different tables and graphs, Method 1 refers to the method that employs only the technique based on the ripple component, and Method 2 the method that combines both sensorless speed techniques.

A. *EMG30 Motor*

The results for different constant speeds are shown in Table. II. Results for a lineal speed change are shown in Fig. 8. The error in this situation is 15.03 r.p.m. for the Method 1 and 0.2 r.p.m. for Method 2. The deviation is 6.81 r.p.m and 8.04 r.p.m. Method 1 produces a small delay, whereas in Method 2 there is no delay. The error related to a speed step is shown in Fig. 9. The settling time for Method 1 is about 0.7 s whereas for Method 2 is 0.3 s.

B. *719E385 Motor*

The results for different constant speeds are shown in Table. III. Results for a lineal speed change are shown in Fig. 10. The error in this situation is 14.24 r.p.m. for the Method 1 and 3.5 r.p.m. for Method 2. The deviation is 8.20 r.p.m and 2.77 r.p.m. Method 1 produces a small delay, whereas in Method 2 there is no delay. The error related to a speed step is shown in Fig. 11. The settling time for Method 1 is about 0.8 s whereas for Method 2 is 0.3 s.

VI. Conclusions

Along this work it has been described a method based on two sensorless speed techniques, based on the dynamic model of the motor and based on the ripple component. In a complementary way it has been also described a method that only employs the technique based on the ripple component. After that, the accuracy of each method has been evaluated in order to see the improvement introduced by the first method.

TABLE II
SPEED MEASUREMENT ERROR FOR MOTOR EMG30

Real Speed (r.p.m.)	Method 1		Method 2	
	Average Error (r.p.m.)	Deviation (r.p.m.)	Average Error (r.p.m.)	Deviation (r.p.m.)
455	454.71	25.50	1.35	2.37
719	127.71	17.82	0.20	2.25
970	7.47	5.43	0.77	4.15
1500	0.06	5.04	0.18	5.14
2000	0.06	5.04	0.53	8.12
2500	0.09	9.75	0.08	9.81
3000	0.73	9.51	0.72	9.48
3500	1.22	12.60	1.31	12.67
4500	1.83	19.14	1.67	18.16

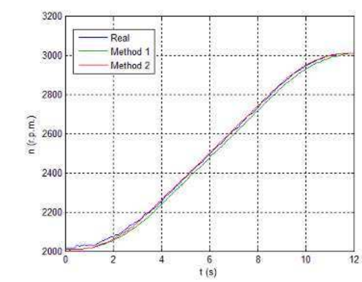


Fig. 8. Lineal variation of speed in motor EMG30.

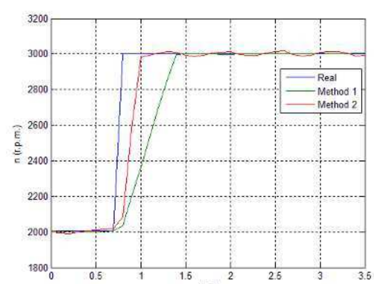


Fig. 9. Speed Step in motor EMG30.

TABLE III
SPEED MEASUREMENT ERROR FOR MOTOR 719E385

Real Speed (r.p.m.)	Method 1		Method 2	
	Average Error (r.p.m.)	Deviation (r.p.m.)	Average Eror (r.p.m.)	Deviation (r.p.m.)
1117	548.51	42.53	24.42	8.35
1290	288.65	13.57	12.40	6.31
1600	152.23	12.03	1.57	6.20
2100	40.63	12.85	0.44	6.23
2578	1.5	7.15	0.4	6.84
3117	1.69	12.12	0.67	11.92
3616	4.28	18.31	3.31	18.49
4115	4.31	11.43	0.15	11.03
4580	9.75	14.41	1.46	13.11
5082	12.04	19.13	2.5	12.94

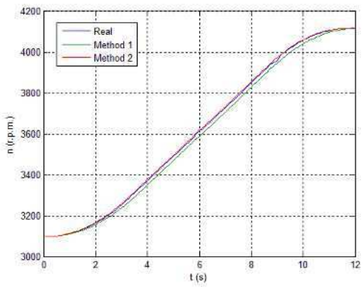


Fig. 11. Lineal variation of speed in motor 719E385.

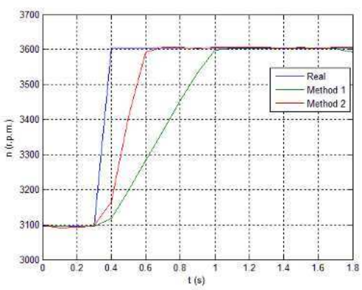


Fig. 10. Speed Step in motor 719E385.

The obtained results for each method after doing the tests with two different dc motors show that for constant speeds the method based on both sensorless speed techniques has greater accuracy at low speed whereas for medium and high speed the accuracy is very similar. When the speed is changing in a linear way (constant acceleration) the method based on both sensorless speed techniques has a better tracking performance, whereas the method based only in the ripple component technique has a small delay of some cents of milliseconds. When a speed step is produced, the settling time is smaller in the method that combines both sensorless speed techniques, approximately the half of the value obtained for the other method.

As consequence of all these results, it is possible to say that the method that combines both sensorless speed techniques provides notable improvements in speed detection, over the method that is based only in the ripple current. It is necessary to say that these improvements require a higher computational cost, because new functional blocks are added to the system. This computational cost does not increase very much the final cost of the system with the current technology level. Therefore, in systems where the speed is small and speed changes are common, it is recommended to use the method based on both sensorless speed techniques because of the introduced improvements. However, if the speed is high and constant the best option will be the method based only in the ripple component.

REFERENCES

[1] P. Vas. “Sensorless vector and direct torque control”. Oxford University Press, 1998.

[2] M. Hilairet, D. Diallo and M. E. H. Benbouzid. “A selfreconfigurable and fault-tolerant induction motor control architecture for hybrid electric vehicles”. Int. Conf. on Electrical Machines (ICEM). September 2006.

[3] M. Hilairet and F. Auger. “Sensorless speed measurement using current harmonic spectral estimation in a DC-motor”. In Proc.SPEEDAM2006 pp 740-745

[4] W. Leonhard. “Control of electrical drives”. Springer Verlag 1985.

[5] P. Chevrel, and S. Siala. “Robust DC-motor speed control without any mechanical sensor”. Proc. Electrimacs, Saint-Nazaire, September 1996.

[6] B. Trump. “DC motor speed controller: control a DC motor without tachometer feedback”. Application Bulletin, Burr-Brown, 1999.

[7] T. Leephakpreeda. “Sensorless DC motor drive via optimal observer-based servo control”. Optim. Control Appl. Methods, 2002, 23, (5), pp. 289–301.

[8] K. Jaszczak, and T. Orlowska-Kowalska,. “Sensorless adaptive fuzzy logic control of DC drive with neural inertia estimator”. J. Electr. Eng., 2003, 3, (1).

[9] J. X. Shen, Z. Q. Zhu, and D. Howe. “Sensorless flux-weakening control of permanent magnet brushless machines using third-harmonic back-EMF”. IEEE Int. Electric Mach. Drives Conf., IEMDC’03, 2003, vol. 2, pp. 1229–1235.

[10] Z. Z. Liu, F. L. Luo, and M. H. Rashid. “Speed nonlinear control of DC motor drive with field weakening”. IEEE Trans. Ind. Appl., 2003, 39, (2), pp. 417–423.

[11] S. R. Bowes, A. Sevinc, and D. Holliday. “New natural observer applied to speed-sensorless DC servo and induction motors”. IEEE Trans. Ind. Electron., 2004, 51, (5), pp. 1025–1032.

[12] S. Li, G. Hai-Hiao, T. Watanabe, and O. Ichinokura. “Sensorless control of DC motors based on extended observers”. 11th Int. Power Electronics Motion Control Conf., EPE-PEMC, 2004, vol. 2, pp. 376–381.

[13] K. Ohishi, Y. Nakamura, Y. Hojo, and H. Kobayashi. “High-performance speed control based on an instantaneous speed observer considering the characteristics of a dc chopper in a low speed range”. Electr. Eng. Jap., 2000, 130, (3), pp. 77–87.

[14] Baoguo Yuan, Zhihua Hu and Zhengxin Zhou. “Expression of Sensorless Speed Estimation in Direct Current Motor with Simplex Lap Winding”, International Conference on Mechatronics and Automation 2007.

[15] S. Erwin Kessler, M. Wolfgang Schulter. “Method for establishing the rotational speed of mechanically commutated d.c. motors”. U. S. Patent, No. 6 144 179. November 2000.

[16] M. John Ma, L. Steven Weiss. “Motor pulse extraction system”. U. S. Patent, No. 4 684 858. August 1987.

Defining Adaptive Assessments Using Open Specifications

Héctor G. Barbosa León, Francisco J. García Peñalvo, María José Rodriguez-Conde
University of Salamanca, Science Faculty, Plaza de los Caidos s/n, Salamanca, Spain 37008, Technological Institute of Colima, Av. Tecnológico s/n, Villa de Álvarez, Colima, México 28976
barbosah@usal.es, fgarcia@usal.es, mjrconde@usal.es

Abstract - **This article presents an authoring tool to construct adaptive assessments covering two aspects: the first one is a proposal to adapt to the learning style of the student by defining links to multimedia material that could be presented to the user and the second one is to adapt the questions presented to the student based on the level of difficulty of each item.**

I. INTRODUCTION

Nowadays the educative content in the Internet is evolving from a static to a dynamic view by adapting the information to the needs or preferences of the users. This content, developed in form of Units of Learning (UoL), include one or several learning objects and their related multimedia material.

One of the activities integrated in those UoL's are the assessments to evaluate the knowledge or achievement of the students. Also, this activity could be seen as an element that allow the entire eLearning system to set a new knowledge level for each student, closing a circular activity and change to new educative content or learning lessons [1](figure 1).

But, traditionally, assessment activity has been seen as a task aside of the e-learning process and could be dangerous to focus the research in assessment specifically [2] as this tends to isolate this activity from teaching and learning in general. For us, the assessment activity after the student had been taken the educative lessons could be seen as an element that contributes to the adaptation of the system to the learner needs or preferences as well to set a new user knowledge level, evaluations and learning profiles [3]. According to the Australian Flexible Learning Framework [4], assessment, especially when is included within a real learning task or exercises, could be an essential part of the learning experience, giving to the entire Web site the characteristic to adapt itself to the needs of the users.

The assessment activity after the student took the educative lesson is an element that contributes to the adaptation of the system to the learner needs or preferences as well [5]. If an item could adapt their presentation to the preference of the user we can reduce the stress because the student may be able to understand the questions more easily and faster. Also, we organize each single item into structures to define another adaptation process using a complexity level classification or "item weight" which is the traditional adaptation process using by some assessment tools [6], [7].

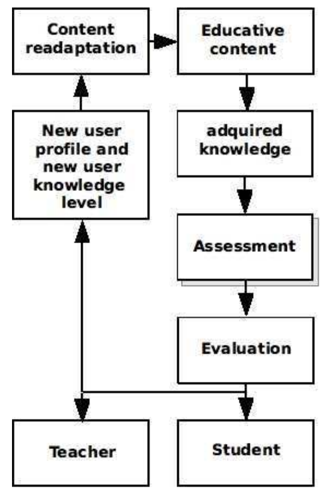


FIGURE 1. ASSESSMENT INSIDE THE LEARNING PROCESS

One consideration to take into account in those tools is the fact that they started as proprietary tools with their own definition for the items and test avoiding sharing their products with other assessment tools. Some of them use open standards and specifications to allow the transference to and from other systems, but this kind of transferences are limited to simple single items, following their definition in the IMS QTI (Instructional Management Systems Question and Test Interoperability) [8].

The rest of this article is organized as follow: in section II we briefly describe the IMS specifications to construct items and exams, in section III we show the adaptation techniques to be applied in our work, in section IV, we describe the adaptation processes and how those are implemented using open specifications, in section V, we describe the products obtained. In section VI we briefly describe the tool to construct adaptive test, in section VII we describe the results obtained and in section VIII we give our conclusions and future work.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_36, © Springer Science+Business Media B.V. 2010

II. DEFINING ASSESSMENT ITEMS AND EXAMS USING IMS SPECIFICATIONS

The IMS CP (Instructional Management Systems Content Package) [9] specification is used when is necessary to transfer learning objects (lessons, exams or other material) between several Learning Management Systems. In this case we can use this specification to package and transfer assessment items between LMS (Learning Management Systems). In the case of a simple element, the package will contain (a) the manifest (XML file called imsmanifest.xml), (b) the item (QTI XML format file) and, (c) any auxiliary files required by the item.

The IMS QTI and IMS LD (Instructional Management Systems Learning Design) [10] specifications allow its integration to define a learning object to evaluate the knowledge acquired by the students when interact with a unit of learning (UoL). The IMS LD includes activities of course instruction, modules, lessons and exams in a generic context (those considered as formative assessment) to support the recent knowledge or to give immediate feedback to the student. However, the IMS specifications are used to define learning objects with extra characteristics like adaptation rules for the final presentation and sequence of the questions. The main structure defined in an IMS LD object is the manifest (figure 2), containing the organization and resources. Inside the organizations section some elements are described, like the roles, properties, activities, environments and methods. The integration of the specifications could be done defining tags and instructions in the *imsld:properties* to control the visibility and order of the elements and the *imsld:conditions* to define decision structures. The environments section is a container for the environment elements, each could be used to describe assessment items for a particular learning style. These structures (the environment ones) could be executed by the LMS in parallel, allowing multiple students of different learning styles to access their own adaptable elements (or an adaptable exam).

But, this work aims to use only the IMS QTI specification to define and construct an adaptive exam, allowing to this kind of object to be integrated and used in more LMS that do not necessarily give support to IMS LD learning objects.

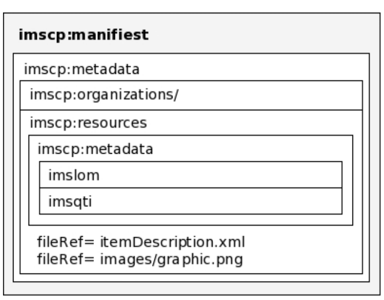


FIGURE 2. IMS LD OBJECT

For this, we designed IMS QTI test objects with adaptation characteristics using bifurcation rules and structures.

III. ADAPTATION TECHNIQUES

Students learn in different ways: seeing and hearing; making reflections and acting; with a logical reasoning process; memorizing and visualizing and by constructing mathematical models. In the same way, the teaching processes vary to make the student to understand better the educative contents and the questions in the summative test. We use the classification of learning styles proposed by Felder [11] and the measurement tool of Index of Learning Styles to obtain the student profile in its preferred learning style.

We can use adaptation techniques and methods [12] to construct adaptive test because we can group items according the associated multimedia material in sections with branching rules to control the access to those sections.

A. Adaptive Presentation

We use the alternative explication because it covers the definition of several versions for the same concept to show to the user according the preferences stored in the user model. About the adaptation technique, we use the page or fragment variants. Using those definitions in the adaptive assessments, we can make equivalence between the versions of the same information fragment with the different items (which use different multimedia material) that look to cover the same instructional objective. We use the same user model that the LMS to select those items that could perform an adaptation to the user's parameters.

B. Adaptive Navigation

The methods to support the adaptive navigation help the user when she/he is using the web page. Those methods are the global guide, local guide, global orientation and the personalized view management. From those, the local guide is the best for us because this method suggests the best next link to follow from the actual node. Talking about the technique, we select the one that hide the links not relevant to the user. When performing a test, we consider the response to the last question to show the next item.

IV. ADAPTIVE PROCESSES

A. Adaptation in the final presentation to the user

The IMS QTI specification allows the developer to define and construct test in which the single items are grouped in several sections which are accessed if some rules are satisfied. We use this specification to adapt the final presentation of each question taking into account the multimedia material referenced from each item.

The QTI specification allows the definition of single questions that include the *ItemBody* section in which we can refer to external multimedia files. After that, we group the items with the same file format into one section (video, audio and text) know as *testPart*.

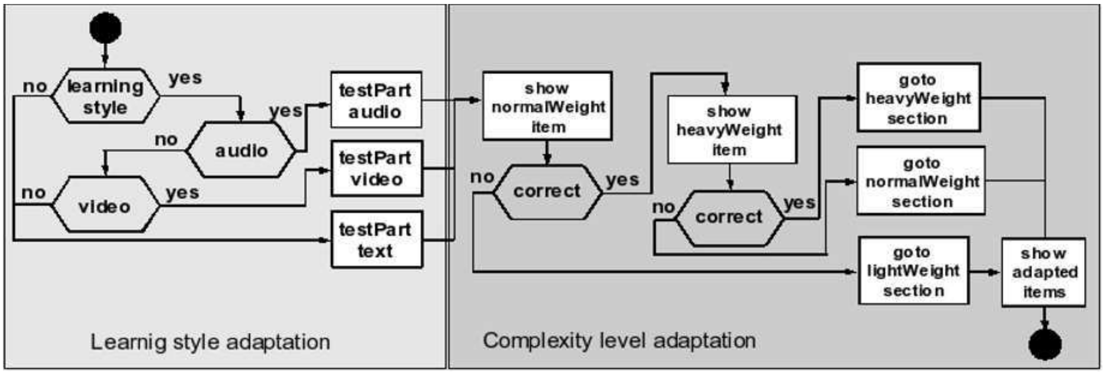


FIGURE 3. PROCESS TO CONSTRUCT AN ADAPTIVE TEST.

B. Adaptation in the Complexity level

This process look to show the next question taking in consideration the result obtained from the user's response to the last question; if the user's response is right then the complexity level of the new question is equal or higher, if the response is wrong then the complexity level is lower than the last question.

With the IMS QTI specification, we can define instructions to control the presentation of a *testPart*, section or single item according to the result of a precondition rule, and with conditional bifurcation to another *testPart*. The precondition and bifurcation labels could be used at any level: *testPart*, *sectionPart*, *assessmentSection* and *assessmentItemRef*. The adaptation process is defined as follow:

- Definition of a testPart for each kind of multimedia material: video, audio and text.
- Definition of some *assessmentSection* inside each *testPart* to group items according their complexity level (high, medium, low), plus an extra medium level item that is showed first to start the adaptation process. With the *preCondition* and *branchRule* labels, we can construct the second process to adapt the test using the complexity level of each question.

The figure 3 shows the construction of test with double adaptation process. First, the learning style of the student is used to select the appropriate *testPart*. If the user model of the LMS does not store a value for the learning style of the student, then the text testPart is chosen. This is the first adaptation process. The second part start with the selection of an item with normal complexity to begin the response-answer process which will select the best next item considering the response and complexity level of the actual question. Inside each testPart, we define three sections containing items of each complexity levels.

V. ADAPTIVE TEST

An adaptive test is a learning object organized in such way it can be used for compatible LMS to give each student with an adapted test using the parameters stored in the user model. The authoring process allows the definition and construction of different products:

- As separate single items that could be referenced by compatible external tools using the IMS QTI specification.
- As complete test that could be used in a session by LMS as a part of a learning activity.
- As a learning object referenced from a learning unit defined using the IMS LD specification in which some conditions are defined to select the testPart using the value stored in the user model.

The elements to perform the adaptation are:

- User preference: stored in the user model with values of "visual", "audio" or "text.
- Result of the first normal-level question stored in the unifiedResult variable, with values of "Y" or "N". This variable is special because look to unify the result for the different kind of questions (simple choice, multiple choice, etc.).
- The result stored in the following questions in the variable unifiedResult, with values "Y" and "N".

In the authoring process the author select from the list, those items that will be included in the test. The tool shows the items grouped by the multimedia files referenced by them.

Once the selection is made, the tool construct the XML-QTI file with the testPart, assessmentSection sections and the reference to the external QTI files for each question and references to the external multimedia files as well. At the end, a file with the IMS CP instruction is generated and packed as .zip format; this file is the one that will be used by the LMS. Also in the authoring process is possible to define the following kind of test:

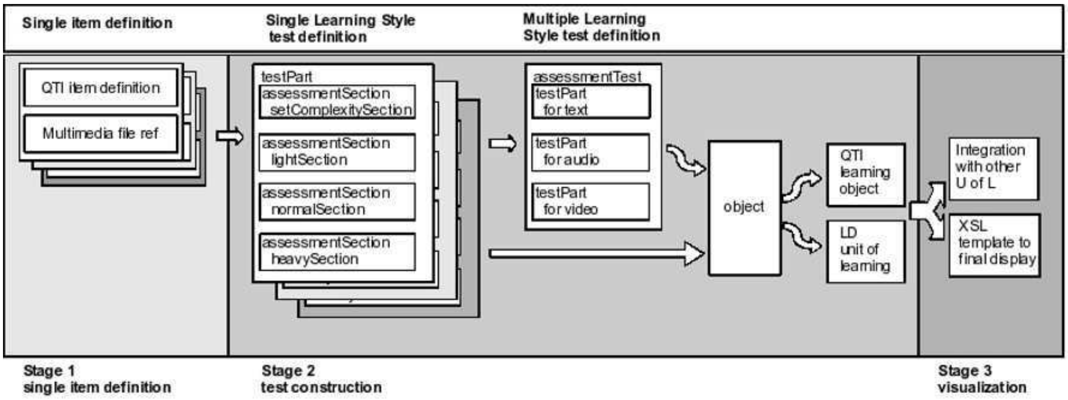


FIGURE 4. CONSTRUCTION PROCESS OF AN ADAPTIVE TEST

- Sequential test: This test does not include branch rules, presenting the questions in the sequential order they were selected from the list.
- Random test: This test does not include branch rules, but the ordering label is set to shuffle.
- Adapted test: This test includes the adaptation rules with the branchRule and preCondition labels to control the items presented to the users.

VI. ADASAT DESCRIPTION

We develop an authoring tool to construct items and test with adaptation characteristics, following open standards and specifications like the IMS QTI, CP and LD and the XML language. The aim is to allow to the author to define items and test using a graphical interface starting from the definition of single questions to the construction of complete test packaged using IMS CP and delivered in a .zip package.

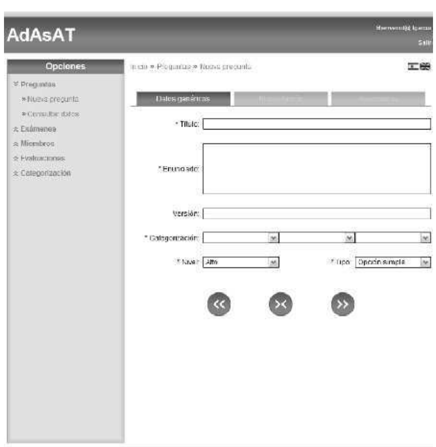


FIGURE 5. CAPTURING A NEW QUESTION

The application covers the following activities:

- Question Construction management.
- Test Construction management.
- User management.
- Test management.

In figure 5, we show the user interface to capture questions. The elements include the title, main statement, question version, categorization and the complexity level of the question, finally, the author select the type of question, including single or multiple option.

VII. RESULTS OBTAINED

To verify that an adapted test is better for the students in terms to ease their answer and understanding, we apply an exam to evaluate the level of knowledge in English language to a group of students in the University of Salamanca. We split this group into two: for the first group we applied the test with questions and accompanying multimedia material that matches their preferences of presentation (audio, text or video); for the second group we applied a test that did not match these preferences. Also, we applied a test a test to determine the learning style of each student, prior to the adaptive test.

Please, recall that this test also perform an adaptation in the level of complexity of each item, for this we organize the items grouping them according this characteristic and, in a higher level, we set a classification according to the multimedia material they are refereeing to.

The aim of this activity is to evaluate our thesis hypothesis that, if a student is presented with an adaptive test that matches its preferences of presentation of the accompanying multimedia material he/she could average better results in a test.

The evaluation included the following statistical test:

Final note: To check if the final note of the test is higher for the students with adapted test.

Time to response: To check if the total time to answer the test is lower when an adapted test is presented to the student.

Level of satisfaction: To check if the level of satisfaction reported by the student is higher when the test is adapted to his/her preferences.

For the final note test the results were: the average final note obtained by the students with an adapted test was 6.6 points, whilst for students without adapted test, the final note was of 5.9 points (t test = 0,836).

In the time of response the results were: the average time for an adapted test was 17 minutes whilst for an not adapted test was 20 minutes (t test=0.964).

The level of satisfaction was obtained from the students themselves from a poll after they answer the test. It cover level of satisfaction from 1 (lower satisfaction) to 5 (highest satisfaction). The results were: for an adapted test 3.9 and for not adapted test 3.6 (t test=0.821).

Bring to mind that, the statistical test were applied only to verify the results for the adaptation process in the presentation of the multimedia material, but the test also perform the adaptation in the complexity level for all participants.

Also, we think that these results must be taken with some considerations at the time to make generalizations. In other words the results obtained in this work may change in other learning environment and stated that this is an internal work. Upon this, we consider that our conclusions are a science contribution for future works in this area.

VIII. CONCLUSIONS

Online assessment is an important step inside the e-learning process, helping to improve the learning and teaching experience.

According to the new developments in the area of e-learning, we can see that most of these developments look to be compliant with accepted standards like the IMS QTI. This gives the convenience to those works to be interoperable and adaptable to several learning platforms. In concordance, referring to the assessment activity, we can think that it must be interoperable as well, because it is one element of the e-learning process and play an important role inside this task.

Adaptability is another key factor in assessment. Given the fact that assessment is an important element of the e-learning process and the fact that this process look to be interoperable, then we can think that the assessment tool could be used with different educative content administrators with different conceptualizations and ways to design an ways of purposes to the students. To face this situation it is necessary to develop a test with different types of resources, different kind of assessments, groups of students, kind of questions, etc.

Under this conceptualization, we created and applied an adaptive test to a group of students to evaluate the hypothesis that, if an assessment tool is adapted to the user`s needs or preferences we can reduce the time to answer the test while we increase the level of satisfaction of the user at the same time.

ACKNOWLEDGMENT

We want to thank to the group KEOPS (ref. TSI2005-00960) of the University of Salamanca for their contributions and ideas for the development of this work. In addition, we want to thank to the Education and Science Ministry of Spain, National Program in Technologies and Services for the Information Society, Ref.: TSI2005- 00960.

Héctor Barbosa thanks the National System of Technological Institutes (SNIT–Mexico) for its financial support.

REFERENCES

[1] Barbosa, H., Garcia F. Importance of the online assessment in the elearning process. 6th International Conference on Information Technology-based Higher Education and Training ITHET, Santo Domingo.

[2] Booth, R., Berwyn, C. The development of quality online assessment in vocational education and training. Australian Flexible Learning Framework, Vol.1, 2003, 17.

[3] Barbosa, H., García F. (2005). A model for online Assessment in adaptive e-learing platform. In proceedings of the 3th. Internatonal Conference on Multimedia and Information and Communication Technologies in Education: m-ICTE 2005, Cáceres, Spain, Vol.l 1, pp. 16-20.

[4] Booth, R., Berwyn, C. The development of quality online assessment in vocational education and training. Australian Flexible Learning Framework, Vol.1, 2003, 17.

[5] Barbosa, H., Garcìa, F., Rodriguez, M. Construction of Assessments with Double Adaptation Processes. CISSE/EIAE 07, University of Bridgeport and the IEEE. Innovations in E-Learning Instruction Technology, Assessment and Engineering, Iskander (ed.), pp. 133-137.

[6] Gouli E., Kornilakis, H., Papanicolau, K., Grigoriadou, M. (2001). Adaptive Assessment Improving Interaction in an Educational Hypermedia System. In procedings of Panhellenic Conference on Human-Computer Interaction, Patras, Greece.

[7] Guzmán, E., Machuca, E., Conejo, R., Libbrecht, P. (2005). LeActiveMath, Integrated Adaptive Assessment Tool.

[8] IMS QTI, Question and Test Interoperability http://www.imsglobal.org/question/index.html. Last accessed 21-03-2008

[9] IMS CP, Content Packaging, http://www.imsglobal.org/content/packaging/index.html. Last accessed 22-03-2008.

[10] IMS LD, Learning Design. http://www.imsglobal.org/learningdesign/index.html. Last accessed 29-03-2008.

[11] Felder, R. M., & Silverman, L. K. (1988). Learning and Teaching Styles in Engineering Education. Engineering Education, 78(7), 674-681.

[12] Brusilovsky, P. 1996. Methods and techniques of adaptive hypermedia. User Modeling and user Adapted Interaction, 6 (2-3), 87-129.

Using the Virtual Model in Teaching Digital Signal Processing

Krystyna Maria Noga
Marcin Radwanski
Gdynia Maritime University, Poland
jagat@am.gdynia.pl,
marrad@gazeta.pl

***Abstract*-These This paper presents a few of Multisim and VisSim use in teaching digital signal processing; it also describes the Internet website containing didactic materials created in the Department of Ship Automation at Gdynia Maritime University. The use of simulation software in education is one of the modern training methods. The package Multisim 8 and VisSim are an easy to use instrument, which enables designing and simulation of a lot of circuits.**

I. INTRODUCTION

The dynamic development of simulation software has resulted in its implementation into traditional didactic methods. The development of technology has caused huge changes in the place and the manner of instructing. The didactic equipment for digital signal processing (DSP), which is used by students for research and simulation during laboratory classes in the Department of Ship Automation at Gdynia Maritime University, is presented in this paper. The laboratory stations are connected together by the internal local area network connected to the Internet. The network has been set up to meet the requirements of the laboratory of DSP. Students' sFTP file-server and the Internet websites [1, 11, 12] are also available; they contain additional didactic materials. Selected virtual projects prepared with Multisim, Matlab, Mathcad, Commsim, and VisSim soft-ware are presented on these websites

Matlab is known for being reliable and fast, but its programming language requires special training. Mathcad is not as fast as Matlab and is less powerful at running large simulation files, but it has significant advantages. The most important is the interface, which is very friendly. Learning to use the interface is much easier than learning a programming language. In higher education the environment of Matlab and Mathcad is well-known and frequently used [2, 3]. Therefore, the paper will present only the examples made with the use of VisSim and Multisim 8. These programs are popular education software, a complete system design tool that offers a large number of components in libraries, a schematic entry, full analog and digital SPICE simulation and radio-frequency capabilities [4]. Multisim and VisSim offer an easy to use graphical interface for all designing and analysis needs. Virtual measuring instruments present simulation results in a graphic mode. The received data can be stored in graphic or text data files. There is a possibility to export the results to Microsoft Excel, Mathcad, Matlab or LabView software. Multisim offers such instruments as: oscilloscope, word generator, signal generator, logic analyzer, Bode plotter, watt-meter, multimeter, network analyzer, distortion analyzer, spectrum analyzer, etc. The way these instruments are used and what they show is very close to the reality. VisSim includes, for example, passive and active low-pass, band-pass, and high-pass filters, Fourier theory, frequency conversion using a mixer, carrier generators, blocks of digital communication using pulse-code modulation, delta modulation, time division multiplexing, FSK and error detection. VisSim enables also the presentation of such problems as: estimation, coding, decoding, modeling transmission channels, modulation and demodulation. To view the results of simulation runs, students have several choices of: a frequency domain, a time domain, XY plots, a log scale, an eye diagram and power spectra.

Computer-simulated experiments using Multisim and VisSim provide a means of reinforcing the important concepts in digital signal processing and electronic communication without the high cost of expensive electronic communication equipment.

II. LABORATORY OF DIGITAL SIGNAL PROCESSING

Recently there have been created many programs that enable designing the systems based on digital signal processing. The examples of such tools are VisSim created by Visual Solutions Inc., and Multisim 8 – by National Instruments. In the Faculty of Marine Electrical Engineering at Gdynia Maritime University, there are two special fields of education: Computer Control Systems and Marine Electro-Automation, where a DSP course is realized in the form of lectures, projects and laboratory sessions. Students carry out a wide range of tasks within laboratory classes. These classes include such problems as: the description of random signals and processes, the Fourier Transform (FT), the Discrete Fourier Transform (DFT), the Fast Fourier Transform (FFT), sampling and reconstruction of signals, the design of digital filters, transformation of images and speech, analogue and digital modulations and modeling communications channels.

Students specializing in Computer Control Systems have in their curriculum 45 hours of lectures (3rd and 4th year of studies), 30 hours of laboratory classes (3rd and 4th year) and 30 hours of projects (4th year). On the other hand, students specializing in Marine Electro-Automation have 15 hours of lectures,

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_37, © Springer Science+Business Media B.V. 2010

15 hours of laboratory classes and 30 hours of projects during 4th year of studies.

Before the participation in laboratory classes, students should prepare projects that depend on the realized subject. Students complete laboratory tasks with the use of real devices and in the virtual environment of Multisim, Matlab, Mathcad, VisSim and Commsim. The software installed on the laboratory server sFTP and HTTP allows students to send their reports in the electronic form directly from home, using any kind of the Internet connection and then to test their projects directly in the computer network. Students can make use of many didactic materials available on the website [1, 11, 12]. This website offers also the discussion forum enabling students to exchange their own experience and to give the information about the realized projects and possible problems.

III. Examples Of Virtual Models

A. *The Fourier Analysis*

During the classes of Digital Signal Processing students carry out, among other tasks, the Fourier analysis of some selected signals. The Fourier transform made the theory applicable to a variety of disciplines, such as signal processing, where the Fourier analysis is now a major tool [5, 6]. The Fourier analysis is also used in image processing, physics and a thermal analysis. Moreover, it is a very important tool in communication [7, 8, 9]. When we send an information signal over a medium, the medium corrupts this signal. The receiver makes a spectral analysis of the received signal and extracts the information. The Fourier analysis allows this. Fourier noticed that we can create any optional signal by summing up simple sine and cosine waves.

The Fourier analysis can be carried out with the use, among others, of Multisim. For example, the presentation of the Fourier analysis of a rectangular signal has been conducted. The results are obtained on the basis of the information concerning harmonic components; then the rectangular waveform is approximated with the use of sinusoidal components. In order to present the Fourier analysis the system presented in Fig. 1 was constructed.

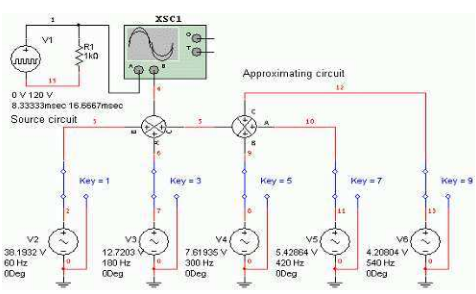


Fig. 1. The diagram of the system tested by the Fourier analysis and the approximation of the rectangular waveform with the use of harmonic components sum.

The examined source was fixed at the voltage of 60 V and the frequency of 60 Hz. The Fourier analysis taking into account 9 harmonics was carried out for the system. The results of the approximation can be observed on the oscilloscope. A similar analysis can be made in the VisSim environment. Among others, on the website [1] there are projects, which make it possible to choose a kind of the signal, i.e. rectangular, sawtooth and triangle waveforms (with the possibility to set its frequency and amplitude) and also to observe the approximation of this signal. The project consists of a series of subcircuits generating the model waveform and the sum of particular harmonics. The sawtooth waveform and its approximation by the Fourier series are presented in Fig. 2.

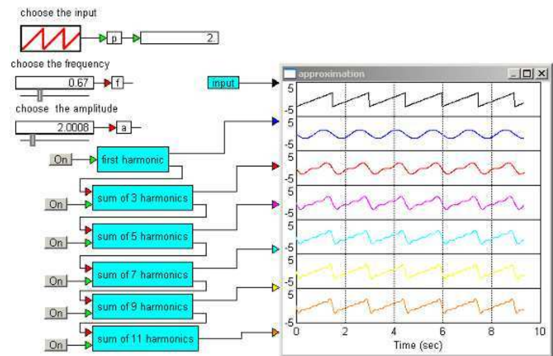


Fig. 2. The sawtooth waveform and its approximation by the Fourier series.

B. *Modulation and Demodulation*

In the systems of digital signal processing the essential role is played by modulation and demodulation. Therefore, the websites [1, 11] present these problems in a very detailed way. Many types of modulation techniques have been devised to represent an information signal [5, 6, 7, 9, 10]. Modulation techniques can be subdivided into two basic groups: analogue modulation and digital modulation. Within analogue modulation the VisSim environment enables students to get to know: Amplitude Modulation (AM), Phase Modulation (PM) and Frequency Modulation (FM). Whereas, in digital modulation it is possible to get to know, among others: Phase Shift Keying (PSK), Quadrature Amplitude Modulation (QAM), Frequency Shift Keying (FSK), Pulse Position Modulation (PPM), Minimum Shift Keying (MSK). On the mentioned websites [1, 11] there are also many projects available containing the examples of modulation. Fig. 3 shows two ways of realizing FSK modulation in the VisSim environment, i.e. with the use of a “ready block” (from the library) representing FSK modulator and the user’s model with two generators of sinusoidal waveforms at different frequencies. Switching the waveforms is realized by the block of the multiplexer Case. In the generators of sinusoidal waveforms the frequency, the amplitude and the initial phase are defined.

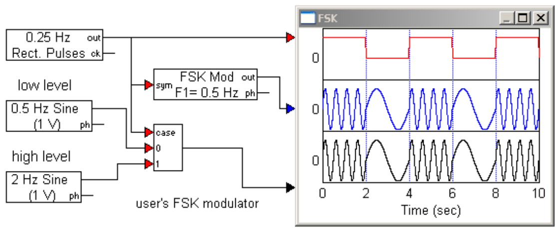


Fig. 3 Virtual FSK modulator in the VisSim environment.

FSK modulator built in the Multisim environment and signal time traces on input and output are shown in Fig. 4.

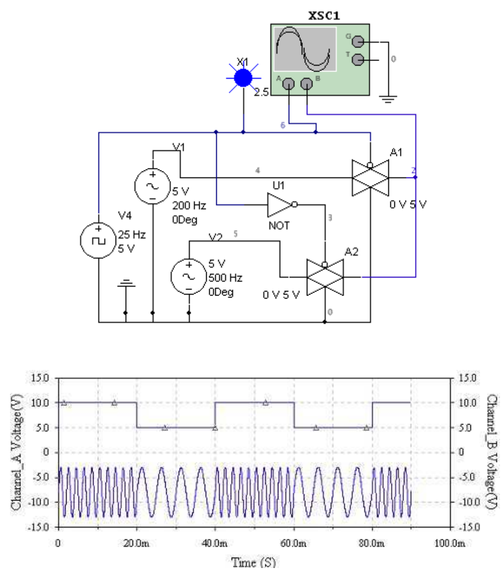


Fig. 4. Virtual FSK modulator in the Multisim environment and time traces of signals.

The main task of the modulation process is to adjust frequency spectrum of a signal to the features of the transmitting channel. Usually the signal containing information includes low frequency components, but transmitting channels have band-pass features. The purpose of a modulation process is to shift the spectrum of the signal into higher frequencies. It also gives more effective broadcasting in case of wireless transmission. During digital signal processing lessons students become familiar with different methods of modulation: analog (AM-SC, AM, SSB-SC, SSB, PM, FM), digital (ASK, FSK, PSK, QAM) and pulse (PAM, PPM, PDM, PCM) [3, 4]. The selection of the appropriate method depends on the bandwidth of the transmitting channel and disturbances in this channel. In the presented example, high speed transmission through a narrow band channel is possible in case of the use of higher-order PSK modulation (QPSK, 8-PSK, 16-PSK). Transmitting signals without higher-order modulation is possible in case of wide bandwidth channels with the use of BPSK or BFSK modulation.

An example of ASK modulation and demodulation prepared in the Multsim environment is shown in Fig. 5. Time traces of signals collected by the virtual oscilloscope are shown in Fig. 6 in the following order: input signal, modulated signal, Schmitt trigger input signal and demodulated signal.

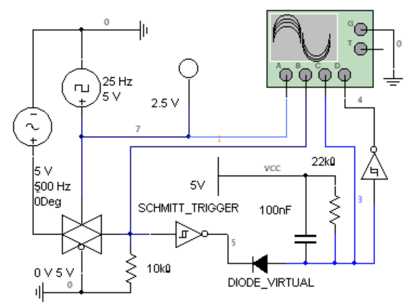


Fig. 5. Virtual ASK modulator in the Multisim environment.

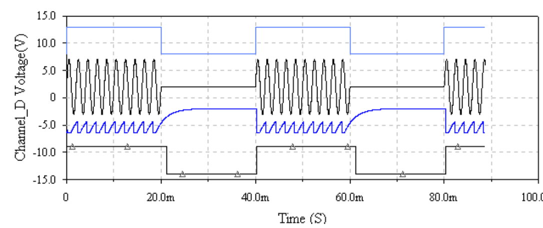


Fig. 6. Signal time traces in the virtual ASK modulator.

ASK modulation isn't used for data transmission because the signal isn't resistant to damping and it causes the deterioration of reception. Its advantage is shutting down the transmitter during the emission of the zero signal. Good resistance to disturbances is characteristic for frequency modulation.

The exemplary simulation of FM modulation and demodulation in the VisSim environment is shown in Fig. 7. The demodulation process was developed in two different ways: using a demodulator block from the VisSim library and by creating the user's demodulator with the use of basic blocks. The output signals from both the blocks are nearly the same, satisfying the expectations. This project shows a possibility of creating specific blocks from the basic components. It is a very valuable feature for educational purposes. The student has an opportunity to become familiar with the internal structure and principles of operation of devices and to design his/her own circuit with the same parameters.

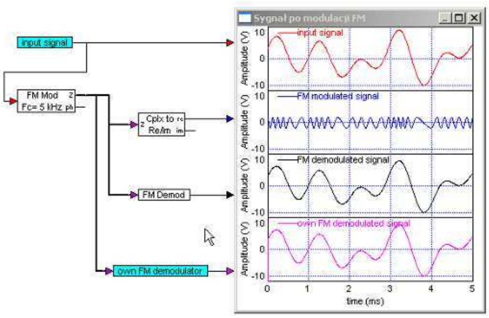


Fig. 7. Virtual FM modulator and demodulator in the VisSim environment and time traces of signals.

C. *Filter Design*

Filters are a kind of circuit very often used in real devices. They fulfill many functions and have many applications. The most common is passing or damping signals at specific frequencies or harmonic components.

Passive filters consist of only such components as: resistors, coils and capacitors, the synthesis and analysis of which is easy to conduct. Multisim software allows the exact analysis of this kind of circuits using proper virtual measuring instruments. The most useful instrument during the filter analysis and simulation is the Bode plotter which calculates and shows a frequency response of the circuit. The example of a simple passive band-pass RC filter with the connected Bode plotter is shown in Fig. 8. The results obtained during the simulation are shown as gain and phase traces (Fig. 9). It can be read from the plots that a given filter passes the signals of the frequency 1 kHz adding a minimal phase shift. The parameters of the virtual instrument, such as bandwidth and resolution, are automatically calculated by the software or can be entered manually by the user.

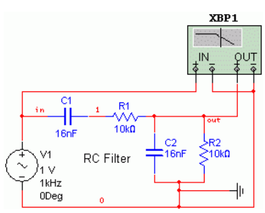


Fig. 8. Circuit with a passive band-pass filter.

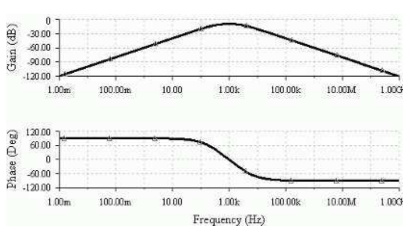


Fig. 9. Results of the analysis shown by the Bode plotter.

The student's task is to identify the type and basic parameters of a filter given by the teacher or to design a filter circuit meeting the requirements.

Active filters are designed with ready templates and equations describing their parameters. The analysis of this kind of circuits is not an easy task. Multisim software allows the easy to conduct simulation and the analysis of these circuits with the use of virtual instruments. The example of a low-pass active Butterworth filter containing the ideal operational amplifier is given in Fig. 10. On the input of the circuit the sum of three signals at different frequencies is put. The sources can be connected with or disconnected from the circuit during the simulation by manual switches. All input and output signals are connected to the virtual oscilloscope and are visible during the simulation (Fig. 11). The output signal is plotted with the bold line. Additionally, the Bode plotter is connected to the input and the output of the filter to obtain a frequency response.

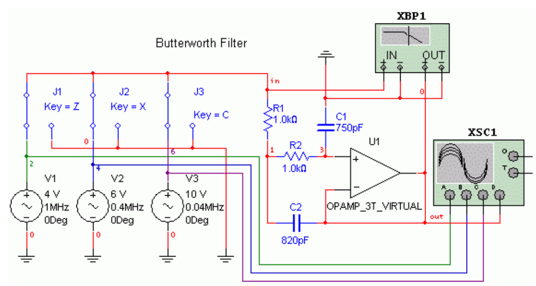


Fig. 10. Active Butterworth filter.

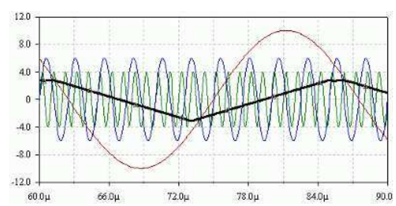


Fig. 11. Signal traces in the circuit with an active Butterworth filter.

Time plots on the oscilloscope show that only the input signal at the lowest frequency passes through the filter. The cutoff frequency of this filter can be read on the Bode plotter screen (Fig. 12). In this case the cutoff frequency is equal to 100 kHz and the phase shift is about -23°.

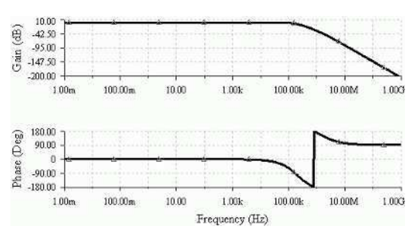


Fig. 12. Gain and phase plots of the Butterworth filter.

In the real world signals contain many components like usable data, disturbances, reflections, etc. Multisim software also enables a spectrum analysis. The example of a simple analog circuit multiplying two harmonic signals with the virtual spectrum analyzer and the oscilloscope connected to the output is shown in Fig. 13. An output complex harmonic signal can be observed on the oscilloscope (Fig. 14). The results shown by the spectrum analyzer (Fig. 15) indicate that the output signal consists of two harmonic components at frequencies 400 kHz and 2 MHz with the amplitude of 40 V.

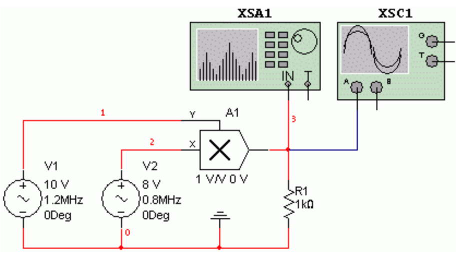


Fig. 13. Simple multiplying circuit.

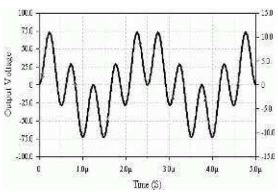


Fig. 14. The output signal of the multiplying circuit.

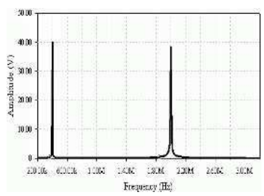


Fig. 15. Spectrum of the output signal.

IV. SIMULATION OF COMMUNICATION CHANNELS

Another significant problem in DSP is data transmission, from this it follows presence of communication channels. The communication channel is one of significant components of the telecommunication system. VisSim also allows students to simulate communication networks in the classroom laboratory. The fundamental elements of any communication system, i.e. the transmitter, the receiver and the channel, can be easily simulated. Students can learn all today's most important communication systems. Channels are modeled with the following blocks: additive white Gaussian noise (complex & real), binary symmetric channel, multipath channel, Jakes channel, Rice and Rayleigh channel, Rummer channel. Signal variability in mobile radio communication environments has been lately extensively studied. Several types of statistical models have been proposed in the literature to describe fading, such as the well-known Rayleigh, Rice, Nakagami-m, Hoyt, Suzuki, Weibull, lognormal [9, 10]. The statistics of the received signal in the mobile radio environment is obtained with the use of the scattering model and the assumption of a large number of randomly phased components.

For example the Nakagami-m channel is presented in Fig. 16. In that channel we assume that the received signal contains an arbitrary number of multipath components. The received signal envelope R can be written as in

$$R = \sqrt{\sum_{i=1}^{m}\left(X_i^2 + Y_i^2\right)} \qquad (1)$$

where X_i and Y_i are an independent Gaussian process, with the mean value equal to zero. The Nakagami-m model is a general fading model, which includes the Rice, the one-sided Gaussian and Rayleigh (for m=1) model. Figure 16 shows a sample of the Nakagami-m fading envelope for m=2 and its histogram, using the VisSim environment. The Nakagami-m fading channel model is commonly used in the mobile communication industry.

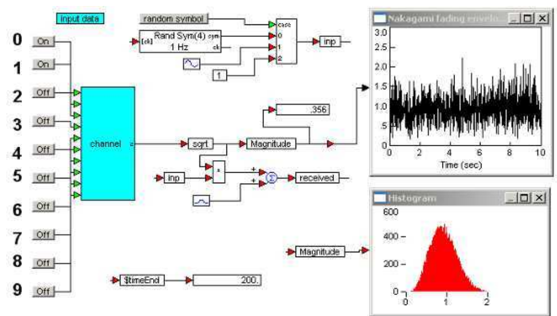


Fig. 16. Nakagami-m model channel, the fading envelope and its histogram.

Another model is presented in Fig. 17. This model makes it possible to choose the following non-selective frequency (flat) fading:

- Rayleigh fading, which is frequently used to model tropospheric or ionospheric scattering in the radio propagation environment with no direct line-of sight (LOS) path, agrees well with the experimental data for mobile systems without LOS [10],
- Hoyt fading is applied in satellite communication and is used when the orthogonal components of the fading signal have normal distributions with different variances and zero average values [10],
- Weibull fading is widely applied in radar systems to model the dispersion of the received signals level produced by some types of clutters, in wireless communication the model seems to exhibit a good fit to experimental fading channel measurements, for both indoor and outdoor environments; recently the topic of digital communication over Weibull fading channels has begun to receive some interest [10, 13],
- Nakagami fading (for m=3) is applied to land mobile and indoor mobile multipath propagation, as well to modeling scintillating radio links [10],
- the channel without fading.

The model also makes it possible to choose modulation/demodulation, i.e. ASK or FSK and the input signal, i.e. eight bits, pseudo-random bit sequence, pseudo noise sequence. Figure 17 shows the results of the Rayleigh channel, ASK modulation and the pseudo noise sequence.

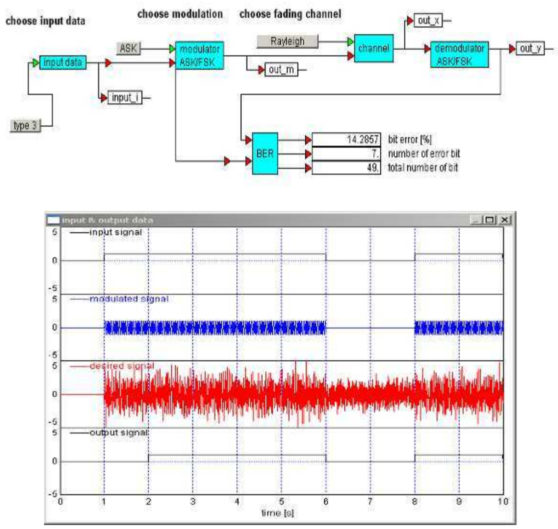


Fig. 17. Virtual communication channel and time traces for the input signal, the modulated ASK signal, the desired signal and the received signal.

V. V. Our Internet Website

Nowadays the Internet access is available and wide-spread all over the world. The interactive Internet websites could be the place not only with the basic information for students, but it could be the place with many other features such as: the examples of reports or the booking system of exercises topics and work deadlines. The website [1] contains the basic information about the laboratory of DSP and the examples of exercises from the range of filters, modulation and encoding made in the Multisim environment.

A wider description of DSP topics is available on the website (in Polish) [11] developed by one of the students. There are: a theoretical description and simulation examples made in the Multisim and VisSim environment from a range of the following topics:

- signal theory (sampling, spectrum analysis, spectral density);
- digital filters (finite impulse response filters – FIR and finite impulse response filters – IIR);
- digital-to-analog and analog-to-digital converters (R-2R, flash converters, successive-approximation, dual-slope, multi-slope, Sigma-Delta)
- the basics of image processing;
- analog and digital modulation methods.

Digital image processing is one of most popular fields of DSP used in practice e.g. in digital cameras. The detailed description with mathematical justification of various methods used in digital image processing is to be found in another student's website [12] (in Polish). This website is created in the e-book style with the examples made in Mathcad software.

VI. Conclusions

The use of computer software in a didactic process allows us to avoid the application of expensive hardware laboratory stations, but not all physical processes can be modeled exactly In many cases the computer simulation simplifies problems and shows only ideal cases. Advanced mathematical modeling can approximate processes to a certain degree, but it may be too difficult for students of the basic DSP course. DSP includes a wide range of problems and the use of virtual models only can be accepted. The hardware implementation of DSP algorithms in digital signal processors or programmable devices should be conducted during an advanced course of telecommunication and related disciplines.

The examples of virtual models presented in this paper are only a small part of the material available on the authors' websites. A large amount of computer simulation software available at present forces universities to choose the best solution for their purposes. The choice depends on particular needs and finances.

References

[1] http://www.am.gdynia.pl/~jagat/
[2] C. Domnisoru, "Using Mathcad in teaching power engineering", IEEE Transactions on Education, Vol. 48, No. 1, February 2005, pp. 157-161.
[3] P. A. Lynn, W. Fuerst, "Digital signal processing with computer applications," John Wiley & Sons, New York, 1994.
[4] "VisSim User Guide" and "Multisim User Guide".
[5] S. Haykin, "Telecommunications systems," WKiL, vol. 1 and 2, Warszawa 1998 (in Polish).
[6] K. Wesolowski, "Basics of digital telecommunications systems," WKiL, Warszawa 2003 (in Polish).
[7] J. G. Proakis, "Digital communication," McGraw & Hill, New York 1995.
[8] J.G. Proakis, D. G. Monolakis, "Digital Signal Processing," Prentice Hall, New Jersey 1996.
[9] B. Sklar, "Digital communications, fundamentals and applications," Prentice Hall P T R, 2001.
[10] M. K. Simon., M. S Alouini, "Digital communication over fading channels", Wiley Interscience, 2005.
[11] http://atol.am.gdynia.pl/tc/cps2007/
[12] http://atol.am.gdynia.pl/tc/Radzienski/
[13] K. M. Noga, G. K. Karagiannidis, "Characterization of Weibull fading channels," *Electronics and Telecommunications Quarterly*, 2007, Vol. 53, No. 1, pp. 47-62.

Comparative Study of Sliding Mode and Relay Enhanced PI Controllers for MIMO Systems

Regina P. and Mija S. J
National Institute of Technology Calicut, Kerala 673601, India

***Abstract*- This paper describes the development of a simple sliding mode controller (SMC) and a relay enhanced multi loop PI controller for multi input multi output (MIMO) processes. In the decentralized SMC design, convergence property of the closed-loop system is guaranteed by means of satisfying a sliding condition. For relay-enhanced design, the control system employs a relay in series with the controller, which yields oscillations for tuning of the PI controller based on an equivalent process model for each loop. Simulation studies are carried out on a two-input-two-output (TITO) process to illustrate the effectiveness of the two methods and to compare their performances. The performance of the controllers is compared also with that of the existing methods.**

I. Introduction

In control applications, most processes have multiple inputs and multiple outputs with interactions in between. The level and density loops of an evaporator, the temperature loops of a multi-zoned furnace, distillation column, robotic manipulators, twin rotors of helicopter etc are good examples of multivariable processes. Lacking general-purpose multivariable controllers, a large percentage of multivariable processes are treated as single variable processes resulting in poor control, wasted energy and materials, inconsistent quality, and plant upsets. Interaction of the control loops is another constraint that has to be considered in the controller design to gain improved real time performance. Most modern multivariable control design methods require a full model of the process. In many cases such a model is not available and physical modeling or system identification may require a prohibitive engineering effort. Furthermore it is hard or impossible, to impose a certain control structure as standard multivariable design methods. Therefore, there is a need for simple methods for tuning controllers for MIMO processes, particularly methods that compromise optimally for engineering efficiency.

A relay in series with the usual PI controllers serves two purposes. First, it gives a high gain, which can yield satisfactory performance even if the controller is not properly tuned. Second, it automatically generates oscillations, based on which an equivalent model of the process can be obtained. With the model, the PI controller can be tuned using one of the many tuning rules now available for PI controllers. It should be mentioned that during the time the relay is used in any particular loop, all other loops are closed, unlike the sequential loop closing method.

A number of designs based on multi loop PID controllers have been reported in the literature [1]. Luyben proposed the biggest log modulus (BLT) tuning method [2] in which, the renowned Ziegler-Nichols rules are modified with the inclusion of a detuning factor, which determines the tradeoff between stability and performance of the system. Individual controllers are designed for the respective loops, ignoring all interactions initially. The calculated controller gains are then scaled by a detuning factor to guarantee stability at the expense of some loss in performance. While only simple computations are involved, the design views interactions as elements impeding system stability and attempts to suppress their effects rather than to control them to speed up individual loops. Hence it is rather conservative. In the sequential loop closing method [3] the loops are closed one after another, with those previously tuned closed with the appropriate controller. The main drawback associated with this method is that the design proceeds in a very *ad hoc* manner. The design decisions made for one of the loops may have undesirable effects on the behavior of the remaining loops. The interactions are well taken care of only if the loops are of considerably different bandwidths and the closing sequence begins from the fastest loop. Generally, the closing of a new loop will bring about some level of interaction to all previously closed loops, thus adversely affecting the overall performance achievable.

The use of relay as an auto tuning aid for a PI controller in the single input single output (SISO) environment was first proposed by Astrom and Hagglund [4], and it has since been successfully field tested on many systems. The limit cycle oscillations induced from a relay feedback experiment provide certain information of the process, which enables the appropriate setting of the parameters of PID controllers. Loh *et al* [5] proposed the most direct extension of the SISO relay auto tuner to a multivariable process based on a combination of both sequential loop closing and relay tuning.

In recent years, there is an increasing interest in the development of robust control systems for processes having uncertainties. Among the available techniques, the sliding mode control strategy appears to be one of the most promising approaches for the control of uncertain processes [6]. David *et al* [7] have systematically analyzed robustness of sliding mode control in the presence of practical engineering constraints and realities. A novel and systematic sliding mode control system design methodology is proposed in this paper. The effectiveness of a variable structure approach to attain desired performance is analyzed and a comparative study is also carried out.

The paper is organized as follows. In Section 2, representation of multi loop control system is discussed. Section 3 deals the design of both relay enhanced PI and SMC controllers. Simulation results are discussed in Section 4 and the conclusions are drawn in Section 5.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_38,

II. EQUIVALENT REPRESENTATION OF MULTI-LOOP CONTROL SYSTEMS

The approach presented in this section is applicable to general *n* x *n* processes [8]. However, the focus is on a 2 x 2 process for illustration of the design principle. The structure of a 2 x 2 multi-loop control system is shown in Fig.1. G_{11}, G_{22}, G_{12}, and G_{21} collectively represents the MIMO process *G(s)*, where G_{12} and G_{21} are transfer functions which represent the inter-loop coupling effects. $G_{c1}(s)$ and $G_{c2}(s)$ constitute the controller for loop 1 and loop 2, respectively. In matrix form, G(s) may be written as

$$G(s)=\begin{bmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{bmatrix} \tag{1}$$

The process is to be controlled in a negative feedback configuration using the multi-loop controller:

$$G_c(s)=\begin{bmatrix} G_{c_1}(s) & 0 \\ 0 & G_{c_2}(s) \end{bmatrix} \tag{2}$$

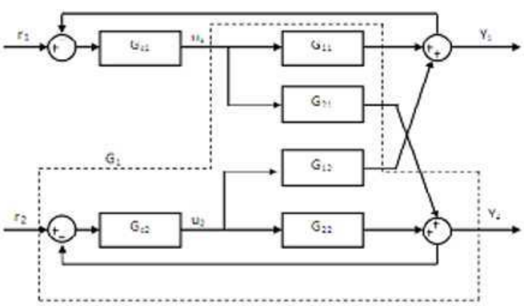


Fig.1. Configuration of multi-loop process

The boxed portion of Fig.1 can be viewed as an equivalent transfer function $G_1(s)$ between input $u_1$and output y_1. It follows that $G_1(s)$ can be obtained as

$$G_1 = G_{11} - \frac{G_{12}G_{21}G_{c_2}}{1+G_{c_2}G_{22}} \tag{3}$$

$$G_2 = G_{22} - \frac{G_{12}G_{21}G_{c_1}}{1+G_{c_1}G_{11}} \tag{4}$$

G_{c1} is the controller for SISO equivalent, G_1 and G_{c2} for G_2. The controllers G_{c1} and G_{c2} are designed with respect to G_1 and G_2, instead of the diagonal elements G_{11} and G_{22}. Many designs of multi-loop controllers are based on the diagonal elements, completely ignoring the off-diagonal terms. In this paper, the cross-coupling effects are considered inherently in the equivalent model as seen by the multi-loop controllers. The n x n multivariable process is modeled by a transfer function of the form (5).

$$\begin{bmatrix} y_1(s) \\ y_2(s) \\ \vdots \\ y_n(s) \end{bmatrix} = \begin{bmatrix} G_{11}(s) & G_{12}(s) & \cdots & G_{1n}(s) \\ G_{21}(s) & G_{22}(s) & \cdots & G_{2n}(s) \\ \vdots & \vdots & \ddots & \vdots \\ G_{n1}(s) & G_{n2}(s) & \cdots & G_{mn}(s) \end{bmatrix} \begin{bmatrix} u_1(s) \\ u_2(s) \\ \vdots \\ u_n(s) \end{bmatrix} \tag{5}$$

$$G_{ij}(s) = \frac{K_{ij}}{\tau_{ij}s \ +1} e^{-\theta_{ij}s} \tag{6}$$

where K, τ and θ are, respectively, the steady state gain, natural period of oscillation and the input delay of the process.

III. DESIGN OF MULTI-LOOP CONTROLLERS

A. *Relay based PI controller*

Under the equivalent representation, the closed loop transfer functions for loop1 and loop2 are

$$G_{y_1r_1}(s) = \frac{G_1(s)G_{c_1}(s)}{1+G_1(s)G_{c_1}(s)} \tag{7}$$

$$G_{y_2r_2}(s) = \frac{G_2(s)G_{c_2}(s)}{1+G_2(s)G_{c_2}(s)} \tag{8}$$

It is assumed that proper pairing of controlled and manipulated variables has been done. If the process is inherently poorly paired, the static relative gain array (RGA) may be used to make possible the necessary arrangement. It is assumed that the inter-loop interaction is not significant and multi-loop controller design can be adequate without a decoupler. Both the controllers G_{c1} and G_{c2} use a PI control structure as given by (9) and (10).

$$G_{c_1}(s) = K_{c_1}\left(1+\frac{1}{T_{I_1}s}\right) \tag{9}$$

$$G_{c_2}(s) = K_{c_2}\left(1+\frac{1}{T_{I_2}s}\right) \tag{10}$$

The extension to a full PID configuration is direct and straightforward by assuming an empirical relationship between T_D and T_I , e.g., the 25% rule: T_D=0.25T_I . Thus the approach is still applicable to PID controllers.

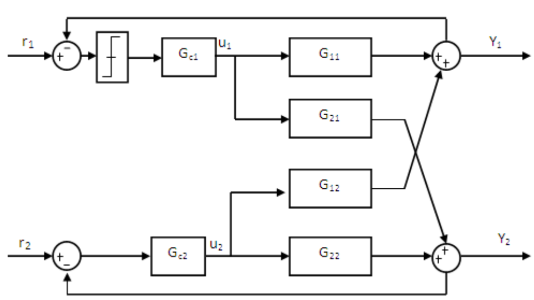


Fig 2. Relay feedback configuration

The relay feedback configuration for the tuning of multi-loop controller is given in Fig.2. Essentially, in this configuration, a relay is inserted into a loop, with all loops (including the one with the relay) already under closed-loop control, although the controllers may not be adequately tuned. Such a configuration is similar to that of the self-oscillating adaptive system. And, it is possible to use the oscillation signals positively to tune the PI controller, and the relay function may be discontinued once the PI controller is tuned.

B. Tuning the multi loop PI controller

The approach is to attempt to fit a low-order transfer function model for the equivalent models $G_1(s)$ and $G_2(s)$, based on the oscillating signals already prevalent in the loop due to the presence of relay. The PI controller is then tuned based on this model. A first-order rational transfer function model with dead time is chosen, given by

$$\widetilde{G}_i(s) = \frac{K_{p_i}}{T_i s + 1} e^{-sL_i} \qquad i = 1,2 \tag{11}$$

This model is simple in structure, with only three model parameters. Yet, it is one of the most common and adequate ones used, especially in the process control industries. The parameters of the model can be estimated from the usual describing function analysis. The describing function of the relay is given by

$$N(a) = \frac{4d}{\pi a} \tag{12}$$

where d is the relay amplitude and a is the amplitude of the limit cycle oscillation. Under the relay feedback, the amplitude (a) and oscillation frequency (ω) of the limit cycle is thus given by the solution of

$$G_i(jw)G_{c_i}(jw) = -\frac{1}{N(a)} \tag{13}$$

The complex equation (13) generates the following two real equations:

$$\left|G_i(jw)G_{c_i}(jw)\right| = \left|\frac{1}{N(a)}\right| \tag{14}$$

$$\arg[G_i(jw)G_{ci}(jw)] + \arg[N(a)] = -\pi \tag{15}$$

Kpi is determined from $K_{pi}=\Delta y_{i,ss} / \Delta u_{i,ss}$ following a change in set point, where $\Delta y_{i,ss}$ and $\Delta u_{i,ss}$ denotes the steady-state change in the output and input of the system, respectively. The remaining two unknown parameters T_i and L_i of the model are obtained from the solution of these equations as

$$T_i = \frac{1}{\omega}\sqrt{\left(\frac{K_{pi} 4 d_i}{\gamma \pi a}\right) - 1} \tag{16}$$

$$L_i = \frac{[\pi + \arg G_{ci}(j\omega) - \tan^{-1} T\omega]}{\omega} \tag{17}$$

$$\gamma = |1 / G_{ci}(j\omega)| \tag{18}$$

Based on this model, many approaches to tune the PI controller for SISO system have been proposed. The PI parameters [9] are given by

$$K_{ci} = \frac{T}{K_{pi}(L_i + T_{ci})} \tag{19}$$

$$T_{I_i} = T_i \tag{20}$$

where T_{ci} is the desired closed-loop time constant.

C. Sliding mode controller

Sliding mode based control configuration with static decouplers for the multivariable process is shown in Fig.3.

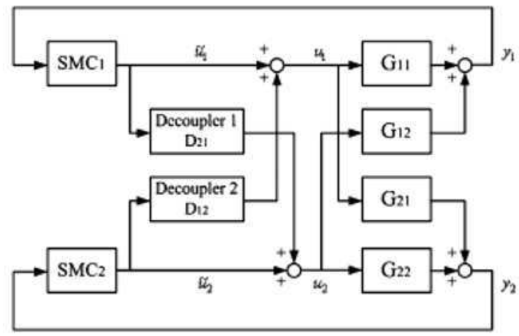


Fig.3. Decentralized SMC configuration

The design of sliding mode controller for MIMO systems is as follows:

The first order plus dead time model is converted into an equivalent state space model as:

$$\begin{aligned}\dot{\widetilde{x}}_1(t) &= \widetilde{x}_2(t) \\ \dot{\widetilde{x}}_2(t) &= -a\ \widetilde{x}_2(t) + b\ u(t-\theta) \\ \widetilde{y}(t) &= \widetilde{x}_2(t)\end{aligned} \tag{21}$$

where, $a = 1/\tau$ and $b = K/\tau$; $\widetilde{x}_1$ and $\widetilde{x}_2$ are the states, and y and u represent, respectively, the model output and the control input. By removing the time-delay from (21), a corresponding delay-ahead prediction model is constructed as follows:

$$\begin{aligned}\dot{x}_1^*(t) &= x_2^*(t) \\ \dot{x}_2^*(t) &= -a x_2^*(t) + b u(t) \\ y^*(t) &= x_2^*(t)\end{aligned} \tag{22}$$

A sliding surface is chosen as:

$$\delta = c_1 x_1^*(t) + c_2 x_2^*(t) = 0 \tag{23}$$

where, c_1 and c_2 are the sliding surface parameters to be designed.

The fundamental idea behind the use of the zero level set of the auxiliary output, denoted by $\Sigma = \{x^* | \delta = 0\}$, as a sliding surface (switching manifold) is to force the controlled motion to adopt Σ as an integrated manifold [10]. When the system trajectory is outside the manifold, the strategy forces the states towards the designed sliding surface. Upon reaching Σ fast switching takes place in the immediate vicinity of Σ, which tries to keep the trajectory constrained to Σ. To eliminate the undesirable switching (chattering phenomena) of the manipulated variable, it is practical to use saturation function instead of the sign function.

$$sat(\alpha/\beta)=\begin{cases}\alpha/\beta, & \text{if}\ |\alpha/\beta|<1\\ \text{sign}(\alpha/\beta), & \text{if}\ |\alpha/\beta|\geq 1\end{cases} \tag{24}$$

where $\beta > 0$ represents the boundary layer thickness. Essentially, the proposed sliding mode controller consists mainly four parameters α, β, c_1 and c_2. The features of these parameters are addressed as follows. The parameter α is related to closed-loop system performance. The larger the value of α, faster the system response at the expense of a larger control input. The parameter β is introduced to eliminate the input chattering. Usually, the selection of β is a tradeoff between control precision and the extent of input chattering. The sliding surface parameters, c_1 and c_2, can be viewed as the weighting factors for the states. Their values can affect the state trajectories and in turn the system performance.

The decentralized SMC system requires *n* separate SMCs. In loop *i* (i = 1,2,. . .,n), the SMC is designed mainly based on the transfer function of $G_{ii}(s)$ and the influences by $G_{ij}(s)$ (j = 1,2,. . .,n, $j\neq i$) through other loops are considered as disturbances that affect this loop. If $G_{ii}(s)$ is a first order plus dead time model, the i^{th} control input is given by

$$u_i(t)=(b_i c_{2i})^{-1}[a_i c_{2i}-c_{1i})x_{2i}(t)-(\alpha_i+sat(\delta_i/\beta_i)] \tag{25}$$

As the multivariable process is diagonal, each of the *n* separate SMCs acts as in the case of single-loop control since loops are independent. In such a case, the stability of the closed-loop system is determined by the stability characteristics of the individual feedback control loops. However, as for interacting multivariable processes the existence of interaction disturbances between loops may affect the stability of the entire multivariable control system to some certain extent. Accordingly, the need has arisen for methods that will provide a decoupling of disturbances caused by interacting loops and thus enhances the stability. Several decoupling methods are available, among them; a practical solution to multivariable process control without the need of a perfect dynamic model is the use of the static decoupling technique. With the static decoupling technique [10], the decoupling gain D_{ij} ($i\neq j$) is given as:

$$D_{ij}=-\lim_{s\to 0}\frac{G_{ij}(s)}{G_{ii}(s)}=-\frac{K_{ij}}{K_{ii}} \tag{26}$$

IV. RESULTS AND DISCUSSION

To verify the effectiveness and applicability of the proposed approaches, both methods are applied to a typical multivariable chemical process. Woods berry eight tray + re-boiler distillation column for separation of methanol and water [8], [10] is the system chosen for illustration.

Design and operation of a distillation column depends on the feed and desired products. The objective is to separate methanol from water. New feed is always being added to the distillation column and products are always being removed. Unless the process is disturbed due to changes in feed, heat, ambient temperature, or condensing, the amount of feed being added and the amount of product being removed are normally equal. Reflux is used to achieve a more complete separation of products. Reflux refers to the portion of the condensed overhead liquid product from a distillation or fractionation tower that is returned to the upper part of the tower as shown in the fig 4.

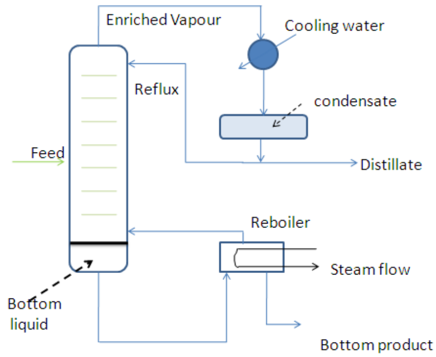


Fig.4. Schematic diagram of distillation column

Wood and berry [8], [10] identified the transfer functions of distillation column considered 2x2 system. The best pair of controlled variable (y_1, y_2) and manipulated variable (u_1, u_2) are decided according to the value of different indices for variable pairing. Reflux (u_1) and vapour boil up (u_2) are the manipulated variables chosen in order to control the distillation composition (y_1) and bottom composition (y_2) in mole fractions. It is not possible to completely purify a mixture of components by distillation. Thus the chosen steady state regimes are y_1=0.96, y_2=0.02.

$$\begin{bmatrix}y_1(s)\\ y_2(s)\end{bmatrix}=\begin{bmatrix}\dfrac{12.8e^{-s}}{16.7s+1} & \dfrac{-18.9e^{-3s}}{21s+1}\\ \dfrac{6.6e^{-7s}}{10.9s+1} & \dfrac{-19.4e^{-3s}}{14.4s+1}\end{bmatrix}\begin{bmatrix}u_1(s)\\ u_2(s)\end{bmatrix} \tag{27}$$

For this particular pairing, the Niederlinski index is 0.4977, Morari resiliency index is 4.06 and the RGA is as follows [1].

$$\Lambda=\begin{bmatrix}2.01 & -1.01\\ -1.01 & 2.01\end{bmatrix} \tag{28}$$

A relay is introduced in series with the initial controller in the first loop; sustained oscillations are introduced in the process by appropriately changing the relay parameters. The critical parameter of the process is identified from the limit cycle oscillations. Amplitude of the limit cycle, *a,* is obtained as 0.8 and 1.6 and the frequency of oscillation, ω is 1.0048 and 0.5024 respectively in loop1 and loop2. From the identified parameters, the controller parameters are calculated and the model of the process is obtained as follows.

$$G_1(s)=\frac{5e^{-1.5878s}}{(9.79s+1)} \tag{29}$$

$$G_2(s)=\frac{10e^{-3.673s}}{(8.76s+1)} \tag{30}$$

Then PI controller parameters are calculated using (19) and (20). To implement the decentralized SMC, the static decouplers gains are set to be d_{21} = -0.34 and d_{12} = -1.48. The parameters for two loops are given in Table1.

TABLE I
PARAMETERS OF THE TWO CONTROL LOOPS

Controller	Loop 1	Loop 2
BLT	$K_c = 0.375$ $K_i = 0.0452$	$K_c = 0.075$ $K_i = 0.0032$
Sequential Method	$K_c = 0.61$ $K_i = 0.065$	$K_c = 0.09$ $K_i = 0.016$
Initial Controller	$K_c = 0.949$ $K_i = 0.1173$	$K_c = -0.131$ $K_i = -0.0145$
Relay enhanced PI Controller	$K_c = 0.2512$ $K_i = 0.0256$	$K_c = -0.0232$ $K_i = -0.0062$
Sliding Mode Controller	$C_1 = 0$, $C_2 = 1.0$ $\alpha = 0.5$, $\beta = 5$	$C_1 = 0$, $C_2 = 1.0$ $\alpha = 0.5$, $\beta = 5$

The MIMO process is simulated in Matlab Simulink for various combinations of loops and controllers. For different PI controllers, two step changes are simulated (u_1=0.96, $u2$=0.02); the first one for loop1 occurring at t=10s and the second one for loop2 occurring at t=120s, to show the interaction between the loops. Fig.5 shows the closed loop performance of loop1 of the MIMO process with BLT, Sequential, Initial PI and Relay enhanced PI controllers. It is clear from the plot that, the relay enhanced PI controller has the best performance in terms of the overshoot, oscillations and settling time.

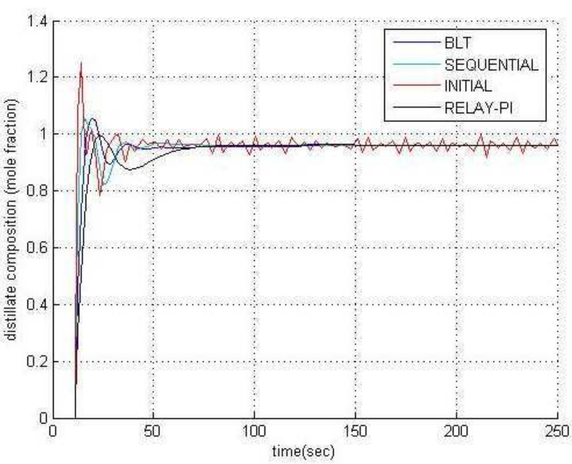


Fig.5. Response of distillation composition for a step change of u_1=0.96, u_2=0.02, with different PI controllers

The closed loop performance of loop2 of the MIMO process with BLT, Sequential, Initial PI and Relay enhanced PI controllers are shown in Fig.6. In this case also relay enhanced PI controller outperforms other controllers.

The effectiveness of the relay enhanced PI controller is more clear from Fig.7, which shows the response of bottom composition for a step change of u_2=0.02 alone.

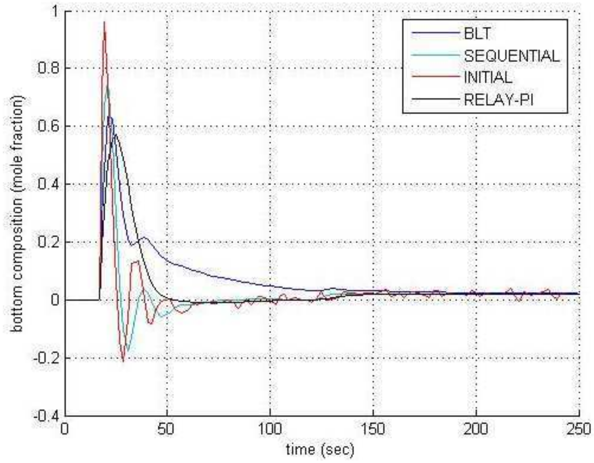


Fig.6. Response of bottom composition for a step change of u_1=0.96, u_2=0.02, with different PI controllers

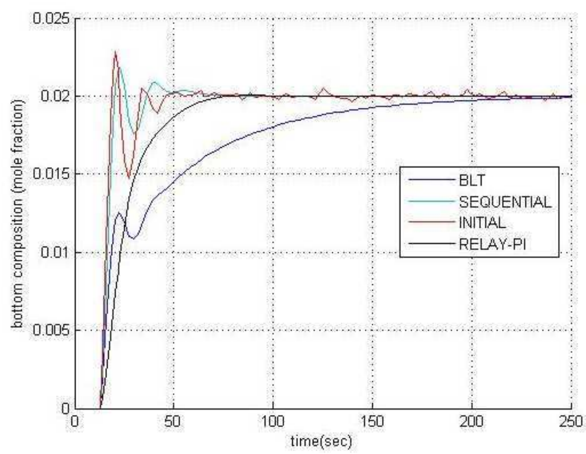


Fig.7. Response of bottom composition for a step change of u_2=0.02 alone, with different PI controllers

In order to compare the performance of SMC and Relay based PI controllers, the system performance is simulated with both the controllers. Improved transient performance is exhibited by the SMC as depicted in Fig.8 and Fig.9. The outputs y_1 and y_2 of loops1 and 2 respectively are plotted against time. The time taken to settle down to chosen steady state regime is much less compared to the PI controllers. There are no oscillations and overshoot is much less for SMC.

To test the robustness of the controllers, simulation study is carried out with forcing a variation of 40% in system parameters. To describe the robustness of the controllers, the bottom mole fraction response in this case is shown in Fig.10. It is clear from the plots that the SMC is very quick in following the nominal response whereas the PI controller fails to follow the nominal response.

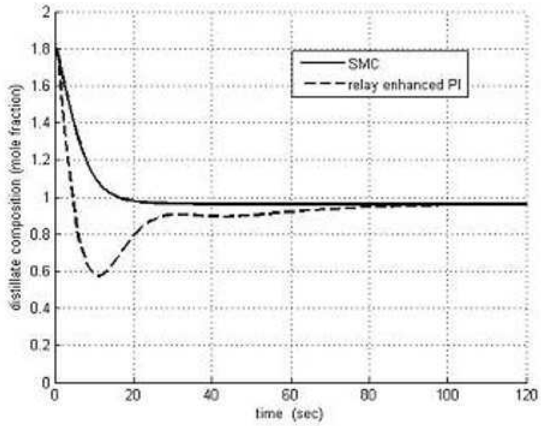


Fig.8. Response of distillate composition with SMC and relay enhanced PI controller

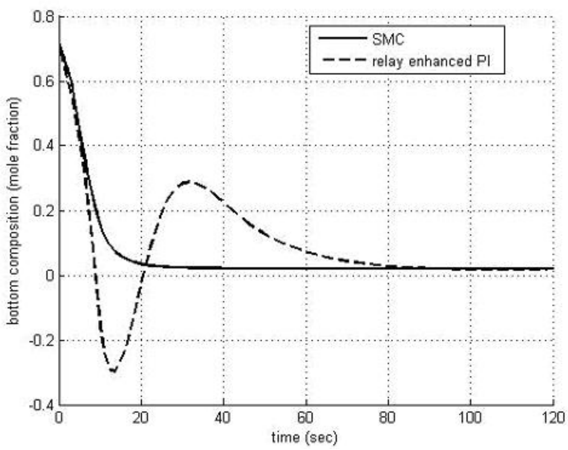


Fig.9. Response of bottom composition with SMC and relay enhanced PI controller

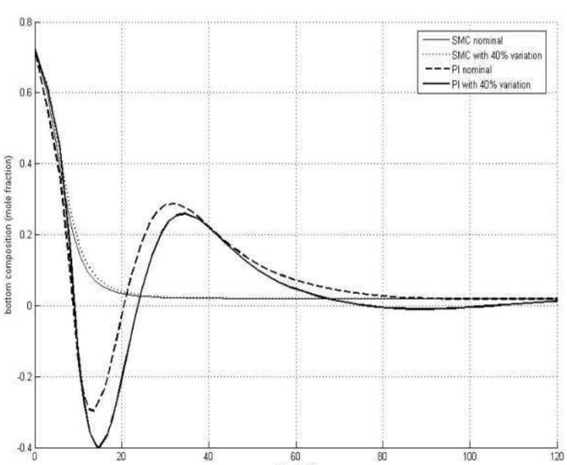


Fig.10. Response of bottom mole fraction for SMC and relay enhanced PI controller with nominal value of parameters and 40% variation

V. CONCLUSIONS

The key feature of the two methods discussed for the control of MIMO processes is that both the methods do not rely completely on the process model. In practical implementation point of view, if more number of states are needed for the controller design, the measurement delays, availability of states and delays in state estimation become critical constraints. In the proposed method for SMC design, since the output alone is required for the computation of control input such delays become insignificant. Due to its significant advantages and good performance, the proposed SMC scheme provides a simple yet robust approach to the control of MIMO processes.

REFERENCES

[1]. W. L. Luyben, "Essentials of Process Control," McGraw-Hill, New York, 1997.

[2]. W. L. Luyben, "A simple method for tuning SISO controllers in a multivariable system," *Ind. Eng. Chem. Process Des. Dev*., vol. 25, pp. 654–660 , 1986

[3]. J. M. Maciehowski, "Multivariable Feedback Design," Addison-Wesley, Workingham, U.K., 1986.

[4]. K. J. Astrom and T. Hagglund, "PID Controllers: Theory, Design and Tuning," *Instrument Society of America*, Research Triangle Park, NC, 1995.

[5]. A. P. Loh, C. C. Hang, C. K. Quek, and V. U. Vasani, "Autotuning of multi loop PI controllers using relay feedback," *Ind. Eng. Chem. Res*., vol. 32, pp. 1174–1180, 1993.

[6]. J. Y. Hung, W. B. Gao and J. C. Hung, "Variable Structure Control: A Survey," *IEEE Trans.Ind.Electron*.,vol. 40, pp. 2-22, 1993.

[7]. K. D. Young, V. I. Utkin and U. Ozguner, "A Control Engineer's Guide to Sliding Mode Control," *IEEE Trans. On Contr. Syst. Tech*., vol. 7, pp. 328-342, 1999.

[8]. K. K. Tan and R. Ferdous , "Relay-enhanced multi-loop PI controllers," *ISA Transactions* vol. 42, pp. 273–277, 2003.

[9]. C. A. Smith, and A. B. Corripio, "Principles and Practice of Automatic Process Control," Wiley, New York, 1997.

[10]. C. T. Chen, S. T. Peng, "Design of a sliding mode control system for chemical processes," *J. of Process Contr.*, vol. 15, pp. 515–530,2005

Modeling, Simulation and Control of Chemical Industrial Reactor

Z. Prokopova, P. Silhavy, R. Silhavy
Tomas Bata University in Zlin,
Nad Stranemi 4511,
760 05 Zlin, Czech Republic
prokopova@fai.utb.cz, psilhavy@fai.utb.cz, rsilhavy@fai.utb.cz

Abstract* – The presented paper is focused on analysis, mathematical modeling, simulation and control of reactor which is used in the chemical and tanning technology. The contribution brings complex analysis of continuous-flow circulation reactor for the cyclohexane production. A mathematical dynamic model was derived and the optimal parameters were computed. Then a comparison of several different methods of controller design was performed. One type of controller was obtained by classical method the second one by algebraic approach via solutions of Diophantine equations in the ring of polynomials. The third type was obtained via solutions of Diophantine equations in the ring of stable and proper rational functions - robust controllers. All simulations were performed in the standard Matlab-Simulink environment.

I. INTRODUCTION

Material and energy balances are the key issues of mathematical models of chemical reactors and processes. The combination with chemical kinetics and transport effects an intellectual basis for chemical reactor design can be obtained. The behavior of real chemical process is obviously too complex for a perfect, complete and whole mathematical analysis. Various kinds of approximations and simplifications are commonly used to obtain results reasonably simple but sufficiently accurate. Cardinal simplifications of the mathematical description can be utilized by the following presumptions:

- Process does not change its properties during the observation period,
- Relation between input and output variables of the process is linear,
- Process variables can be measured continuously and have a continuous-time trend.

A useful and interesting application of reactor design for chemical engineers can be seen in [1], [2]. There are emphases on numerical solutions which are needed for most practical problems in chemical reactor design. Sophisticated numerical techniques and methods are rarely necessary.

The aim of the presented approach is to make the techniques understandable and easily accessible to design engineers and the methodology enables the deeper and comprehensive focus on the chemistry and physics of the process.

There are many reputable authors and superior texts on chemical reactions and theirs kinetics, engineering and reactor analysis and design, e.g. [3], [4], [5], [6].

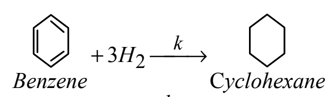
Design of controllers by algebraic approach became a useful and popular discipline in control theory during the last decade. On the beginning the ring of polynomials was used for control design. All controllers are obtained via linear Diophantine equations in an appropriate ring. The necessity of robust control was naturally developed by the situation when the nominal plant (used in control design) differs from the real (perturbed) one. A suitable tool for parameter uncertainty is the infinity norm H_∞. Hence, a polynomial description of transfer functions had to be replaced by another one. A convenient description adopted from [7], [8], [9] is a factorization approach where transfer functions are expressed as a ratio of two Hurwitz stable and proper rational functions (R_s). Then conditions of robust stability can be easily formulated in algebraic parlance.

II. ANALYSIS AND MATHEMATICAL MODEL

Presented reactor is a continuous-flow circulation reactor for the cyclohexane production. Cyclohexane, as an important base material for polyamide manufacturing, makes by catalytic hydrogenation of benzene. A large number of reaction heat transfers during hydrogenation process. Therefore, benzene and cyclohexane mixture hydrogenation proceed in continuous time circulating reactor under hydrogen excess. The presented mathematical model should respect the following conditions:

- Reaction mixture is perfectly mixed by circulation.
- The volume of reaction mixture is constant.
- All technological parameters in reactor are constant.
- Heat transfer on the both sides of wall is ideal.

The hydrogenation mechanism of benzene is supposed to be a first-order reaction:



$$r_B = -k \cdot c_B \qquad (1)$$

where k is a constant of proportionality known as the *rate constant* and r_B is *the rate of the reaction.* There is not the rate at which a particular component reacts. Components benzene and hydrogen are consumed by the reaction and thus are "formed" at a negative rate. The material balance for the

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_39, © Springer Science+Business Media B.V. 2010

component benzene satisfies the following differential equation (for constant volume V of reactor):

$$q_V.c_{BV} = q.c_B + V.k.c_B + V\frac{dc_B}{dt} \tag{2}$$

where q_v and q are input and output flow and c_{BV} and c_B are input and output concentration of component benzene.

The temperature balance in the reactor satisfies the differential equation:

$$\begin{aligned} &q_V.\rho_V.c_{pV}.T_V + (-\Delta H_R)V.k.c_B = \\ &q.\rho.c_p.T + F.\alpha(T - T_x) + V.\rho.c_p\frac{dT}{dt} \end{aligned} \tag{3}$$

where the relation $(-\Delta H_R)V.k.c_B$ is the heat generated by the reaction and $F.\alpha(T - T_x)$ represents the heat transfer into surroundings.

The rate constant k for elementary reactions is expressed by Arrhenius equation:

$$k = k_0 T^n \exp\left(\frac{-E}{RT}\right) \tag{4}$$

where parameter n=0 (or n=0.5, 1 for the special cases of reactions) depends on used theoretical model. E is the *activation energy* and the fraction E/R is called the *activation temperature.*

After some mathematical manipulations the equation (2) of the benzene concentration variation takes the form:

$$\frac{dc_B}{dt} = \frac{q_V}{V}\cdot c_{BV} - \frac{q}{V}\cdot c_B - k_0 e^{-\frac{E}{RT}}\cdot c_B \tag{5}$$

Similarly, variation of the reaction mixture temperature in the reactor is:

$$\begin{aligned} \frac{dT}{dt} = &\frac{q_V}{V}T_V + (-\Delta H_R).\frac{k.c_B}{\rho.c_p} - \frac{q}{V}T \\ &- \frac{F.\alpha}{V.\rho.c_p}(T - T_x) \end{aligned} \tag{6}$$

The time response of the reaction mixture concentration depends on the input flow of the benzene, the concentration of input stream c_{BV}, the volume of reactor filling, the specific heat capacity of reaction mixture c_p and the rate constant k.

The solution of ordinary differential equation (5) requires an initial condition: $c_{BV} = c_{B0}$ at $t=0$.

From all mentioned parameters the input flow stream into the reactor q_v and its initial concentration c_{BV} can be changed.

III. Control Design

Control systems depicted in Fig. 1. and Fig. 2. are considered for design of one degree of freedom – feedback (FB) and two degrees of freedom – feedback feed-forward (FBFW) controllers.

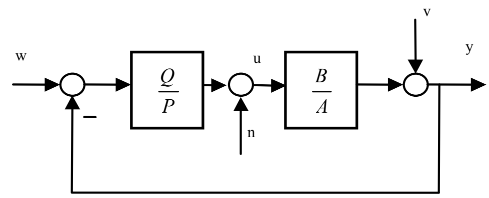


Fig. 1. Feedback control system

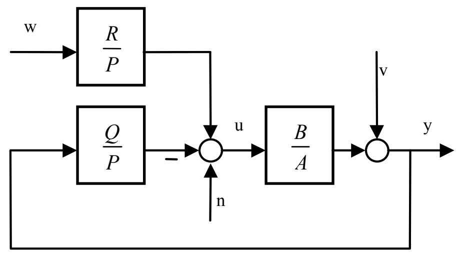


Fig. 2. Feedback - feed-forward control system

Transfer functions $w = \frac{G_w}{F_w}$ and $v = \frac{G_v}{F_v}$ represent the reference and disturbance signal, respectively.

A. Classical approaches

1) Naslin method

The method is determined for the low degrees transfer functions in the form $\frac{B}{A}$. The controller has defined structure in advance. We choose the controller handling that maximal overshooting not exceed limitations. Procedure can be described in the next steps:

-We have the transfer function

$$G(s) = \frac{K}{a_n s^n + \ldots a_1 s + 1} = \frac{B(s)}{A(s)} \tag{7}$$

- The controller has defined structure

$$C(s) = \frac{Q(s)}{P(s)} \tag{8}$$

- We have to calculate the characteristic polynomial

$$AP + BQ = c_0 + c_1 s + \ldots + c_m s^m \tag{9}$$

- For: $c_i^2 \geq \alpha\, c_{i-1} \,.\, c_i+1$, where the relation between α and overshooting is defined by table. For maximal overshooting $|\Delta y| < 5\%$ is $\alpha = 2.0$.

The method provide only parameters borders but their availability is necessary to test by simulation.

2) Method of optimal module

The method was worked out by Oldenburg and Sartorius. The principle is that the amplitude frequency characteristics have to be flat in the area of lowest frequency. The method is very simple for the controlled systems with the transfer function in the form $G(s)=\dfrac{K}{a_n s^n+...a_1 s+a_0}$.

For n=3 and PI controller there is exists explicit formulation of controller parameters:

$$r_{-1}=\frac{1}{2K}\cdot\frac{a_1^2-a_2}{a_1a_2-a_3}\,,\; r_0=\frac{1}{2K}\cdot\frac{a_1^3-2a_1a_2+a_3}{a_1a_2-a_3}\,. \tag{10}$$

B. Modern approaches

1) Pole-placement control

Basic signification in the theory of linear control has the characteristic equation of closed control loop. There are known polynomials of numerator and denominator of controlled system in the characteristic equation but doesn't known polynomials of controller transfer function. The design of controller was reduced for solving Diophantine equations in the modern interpretation, where:

- *a(s) p(s) + b(s) q(s) = c(s)* has solution
  ⇔ Highest common divisor HCD (*a,b*) divide *c*
- If exists $p_p(s)$, $q_p(s)$ equation *a(s) p(s) + b(s) q(s) = c(s)*

$$\begin{aligned} p(s)&=p_p(s)+b(s)t(s)\\ q(s)&=q_p(s)-a(s)t(s)\end{aligned} \tag{11}$$

where *t(s)* is arbitrary polynomial; HCD (*a,b*) = 1

- Particular solution we can obtain by e.g. indefinite coefficient method**.**

Solution:

- Estimation of degrees

$$\begin{aligned} &\left.\begin{aligned}\deg p&=\deg b-1\\ \deg q&=\deg a-1\end{aligned}\right\}\deg c<\deg a+\deg b\\ &\left.\begin{aligned}\deg p&=\deg c-\deg b\\ \deg q&=\deg a-1\end{aligned}\right\}\deg c\ \ \geq\deg a+\deg b\end{aligned} \tag{12}$$

- Calculation of particular solution: By comparison of left hand side and right hand side of Diophantine equation we can obtain set of linear algebraic equations.
- Calculation of all solutions: $\left.\begin{aligned}p(s)&=p_p(s)+b(s)t(s)\\ q(s)&=q_p(s)-a(s)t(s)\end{aligned}\right\}$ *t(s)* is arbitrary polynomial.

2) Robust control

In this case linear transfer functions are no more represented as a ratio of two polynomial but two elements of an another ring. For the purposes of robust control in the sense of this contribution any transfer function *G(s)* of a linear system can be expressed as a ratio of two elements:

$$\begin{aligned} &G(s)=\frac{b(s)}{a(s)}=\frac{B(s)}{A(s)}\; A(s)=\frac{a(s)}{m(s)}; B(s)=\frac{b(s)}{m(s)}\\ &m(s)=(s+m)^n\,; n=\max\{\deg a;\deg b\}\end{aligned} \tag{13}$$

Elements *A, B,...* constitute a sub ring $R_m(s)$ of the ring of all Hurwitz stable and proper rational functions $R_s(s)$. A class of PID like controllers is generated by first and second order systems.

For $m = 0$ subring $R_m(s)$ expands to the traditiona1 ring $R_s(s)$ [7]. An infinity norm (convenient for uncertainty) in both rings is defined by

$$\begin{aligned} &\|G\|=\sup_{\mathrm{Re}s\geq 0}|G(s)|=\sup_{\omega\in E}|G(j\omega)|\\ &\|G_1\,G_2\|=\left\|\begin{matrix}G_1\\ G_2\end{matrix}\right\|\sup_{\mathrm{Re}s\geq 0}\left\{|G_1(s)|^2+|G_2(s)|^2\right\}^{\frac{1}{2}}\end{aligned} \tag{14}$$

This norm is the radius of the smallest circle containing the Nyquist plot of the transfer function. Almost all mathematical models differ from physical systems.

Let $G(s)=\dfrac{B(s)}{A(s)}$ be a nominal plant and consider a family of perturbed systems $G'(s)=\dfrac{B'(s)}{A'(s)}$ where

$$\begin{aligned} &\|A-A'\|\leq\varepsilon_1\ \|B-B'\|\leq\varepsilon_2\\ &\text{or } \|A-A'\ B-B'\|\leq\varepsilon\end{aligned} \tag{15}$$

The control design in $R_m(s)$ is proposed in [11], [12], [13]. For the given nominal plant *G(s)* in the form of (13) it consists of the following steps:

Step 1: All stabilizing (feedback) controllers are given by all solutions of the equation:

$$AP+BQ=1 \tag{16}$$

in the parameter form expressed by the fraction:

$$\frac{Q}{P}=\frac{Q_0-AT}{P_0+BT} \tag{17}$$

where P_0,Q_0 x $R_m(s)$ are particular solutions and *T* is arbitrary in $R_m(s)$.

Step 2: Feedback controllers (if necessary) are given in a similar way by all solutions of

$$F_w S + BR = 1 \tag{18}$$

Step 3: Compute P x $R_m(s)$ such that

F_w divides P for asymptotic tracking problems (19)

F_v divides P for disturbance rejection problems

Step 4: For perturbed plants choose such P, Q in which fulfils the conditions

$$\varepsilon_1\|P_0+BT\|+\varepsilon_2\|Q_0-AT\|\le 1$$
$$\text{or } \varepsilon\left\|\begin{matrix}P_0+BT\\Q_0-AT\end{matrix}\right\|\le 1 \tag{20}$$

Steps 1 - 3 represent the first and simplest case of robust design. All solutions of Diophantine equations in $R_m(s)$ are expressed as functions of the parameter $m > 0$. The va1ue of this parameter strongly influences dynamics as well as robustness of the proposed control system.

IV. Simulation Of Mathematical Model

The continuous-flow circulation reactor for catalytic hydrogenation of benzene with V=0,5 m^3 volume was used for the simulation of the concentration and temperature characteristics. Initial input flow into the reactor was q_v=0.0067 m^{3}s^{-1} with the initial concentration of benzene c_{Bv}=1.9 kg/m^3. Initial rate constant was considered k= 1.616.e14 s^{-1}, density of mass ρ=985 kg m^{-3}, specific heat capacity c_p = 4.050 J/kg.K, transfer heat coefficient E= 435.00 W/m^2.K, surface F=5.5 m^2, activation energy E=4.8e4 J mol^{-1}, initial mass temperature T_v=320 K and temperature of surrounding T_x=290 K.

Steady-State Analysis Results

The reactor steady-state characteristics were obtained by solution of equations (4)-(6). The initial conditions were computed by a standard optimization method - iterative procedure.

Dynamic Analysis Results

For dynamic analysis purposes one input and one output choices were considered. Then were defined input u and output y as deviations from their steady-state values – $u(t) = q(t) - q^s(t)$ and $y(t) = c_B(t) - c_B^s(t)$.

Graphical interpretation of simulation experiments – the output y time responses to input u step changes are depicted on Fig. 3.

Model approximation

For control requirement, described continuous time circulating reactor was approximated by the second order system in the form:

$$H(s)=\frac{1.9}{71s^2+10s+0.5}\approx\frac{B(s)}{A(s)} \tag{21}$$

The approximation was obtained by a four-parameter method adopted from [13] applied to step responses.

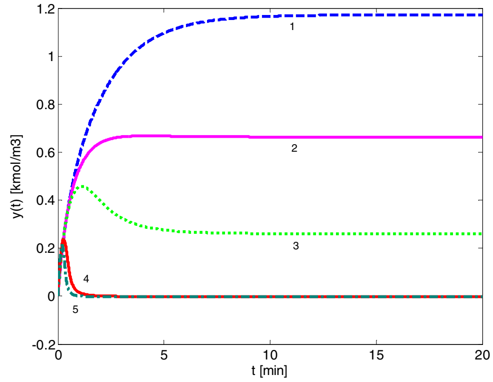


Fig. 3. The output y time responses to input u step changes, u=-0.08(*1*), -0.07(*2*), -0.05(*3*), 0.05(*4*), 0.08(*5*)

V. Results Of Controlled Systems Simulation

One of the main problems of the catalytic hydrogenation of benzene in continuous-flow circulation reactor is progressive degeneration of reaction surfactant. This is cause that the reaction is slowing down, the rate constant k is decreasing and the concentration of benzene in output mixture is increasing. The aim is to keep the concentration of benzene under the critical value. One of the possibilities how to solve the problem is reduction of input flow q of the benzene and cyclohexane mixture into the reactor.

A. Classical control algorithms

Simulation 1: Naslin method, maximal overshooting 5% (α = 2.0), set point w=0.002.

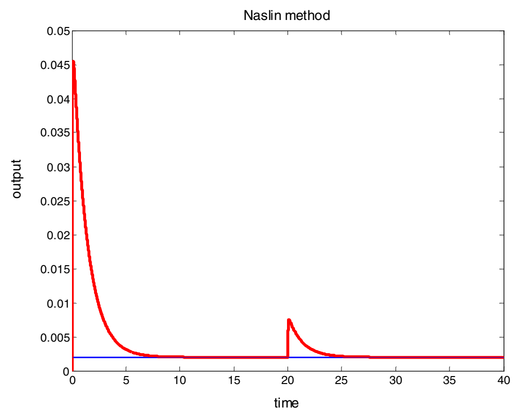


Fig. 4. Results of Naslin method

<u>*Simulation 2:*</u> Method of optimal module, set point w=0.002.

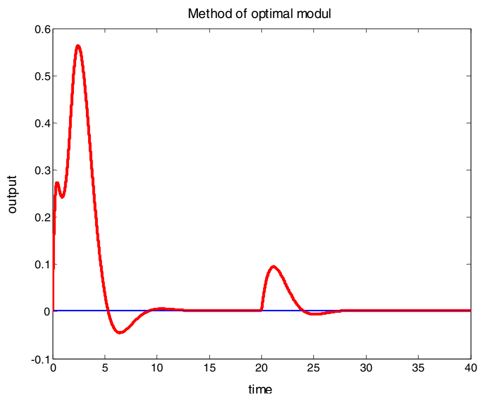


Fig. 5. Results of Method of optimal module

B. Modern control algorithms

<u>*Simulation 3:*</u> Polynomial approach, set point w=0.002.

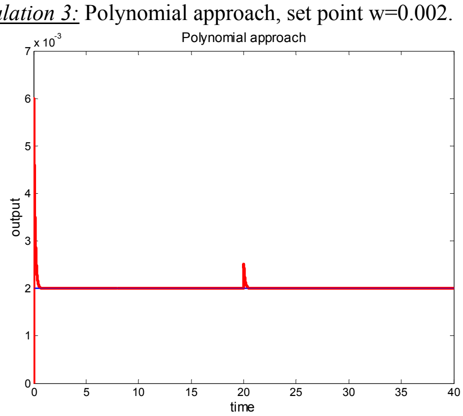


Fig. 6. Results of polynomial approach

<u>*Simulation 4:*</u> Robust approach, FB, m=2, set point w=0.002.

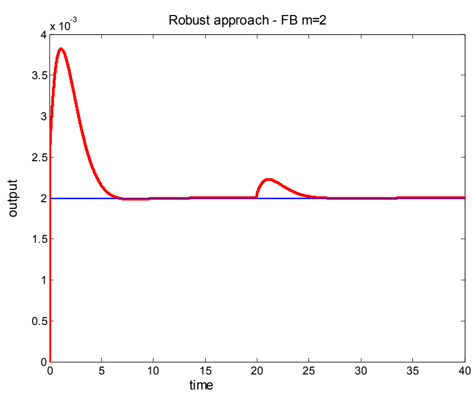


Fig. 7. Results of robust approach – FB, m=2

<u>*Simulation 5:*</u> Robust approach, FB, m=10, set point w=0.002.

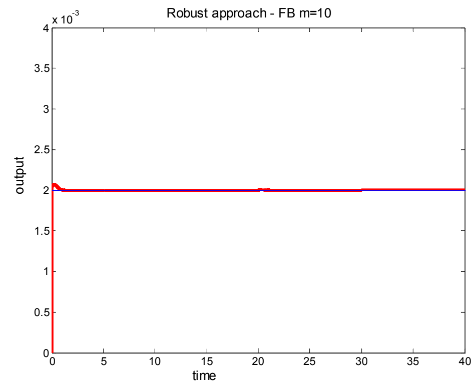


Fig. 8. Results of robust approach – FB, m=10

<u>*Simulation 6:*</u> Robust approach, FBFW, m=2, set point w=0.002.

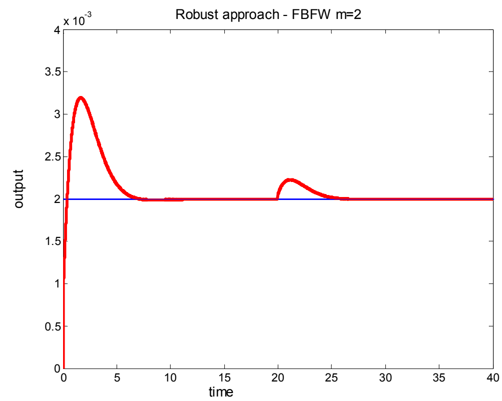


Fig. 9. Results of robust approach – FBFW, m=2

<u>*Simulation 7:*</u> Robust approach, FBFW, m=10, set point w=0.002.

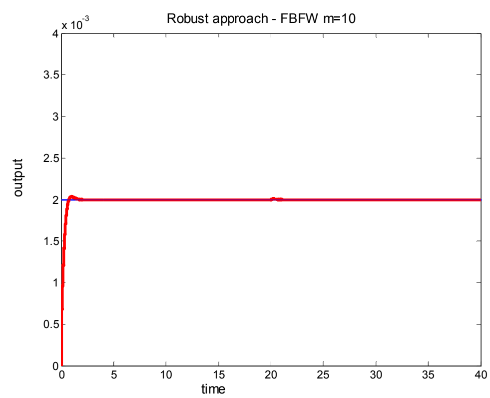


Fig. 10. Results of robust approach – FBFW, m=10

Conclusion

There are several different methodologies presented in the paper. The main aim was to compare classical and modern approaches to control design of one a bit special technological process. Controllers obtained by classical methods have an advantage in simple deduction of controller's parameters but control behaviors from the system have relatively big overshooting. The polynomial approach is much better, but the calculations of controller's parameters are more complicated and controlled outputs are very dependent on the approximation model, which we used for controller design. The robust methodology enables to tune and influence the robustness and control behavior by a single scalar positive parameter *m*. It means that control output is not so dependent on the approximation model of the system. We can see (Fig.7 – Fig.10) that the control behavior is improving with the increasing of parameter *m* value. For *m*=2 is overshooting bigger than we are able to accept but for *m*=10 is overshooting minimal. The control behavior is a little bit better when we use two degree of freedom control loop (feedback and feedforward control system).

The design methods were developed for SISO systems generally. All computations and simulation were performed in the MATLAB+SIMULINK environment.

Acknowledgment

This research has been supported by the Ministry of Education of the Czech Republic in the range of the project No. MSM 7088352102. This support is very gratefully acknowledged.

References

[1] F. Froment, and K. B. Bischotf, *Chemical Reactor Analysis and Design*, 2nd ed., Wiley, New York. 1990.

[2] M. B. King, and M. B. Winterbottom, *Reactor Design For Chemical Engineers*, Chapman & Hall, London. 1998.

[3] R. M. Felder, R. W. Rousseau, *Elementary Principles of Chemical Processes*, 3rd ed., Wiley, New York. 2000.

[4] O. Levenspiel, *Chemical Reaction Engineering*. 3rd ed., Wiley, New York. 1998.

[5] L. D. Schmidt, *The Engineering of Chemical Reactions*, Oxford University Press, New York. 1998.

[6] J. M. Smith, *Chemical Engineering Kinetics*. McGRAW-HILL, New York. p. 131. 1956.

[7] M. Vidyasagar, *Control System Synthesis: A Factorization Approach.* MIT Press, Cambridge , 1985.

[8] J.C. Doyle, B.A. Francis, A.R. Tannenbaum. *Feedback Control Theory*. Maxwell McMillan, 1992.

[9] V. Kučera, "Diophantine Equations in Control – A Survey", *Automatica,* 29, No.6, pp. 1361-75. 1993.

[10] M. Morari, E. Zafiriou, *Robust Process Control*. Prentice Hall, New Jersey. 1989.

[11] R. Prokop, J.P. Corriou. "Design and Analysis of Simple Robust Controllers". *Int. JournalofControl,* Vol. 66, No. 6, pp. 905-921. 1997.

[12] R. Prokop, P. Dostál. "Design of Simple Robust Controllers". In *10th International Conference on PC'95*. Tatranské Matliare, pp. 103-108. 1995.

[13] K.J., Astrom, T. Hagg1und, PID Controllers: *Theory, Design and Tuning*, 343 p. Instrument Society of America. 1995.

A Robust Method for Detecting Vehicle Positions and Their Movements Even in Bad Weather Using Infrared Thermal Images

Yoichiro Iwasaki and Shinya Kawata
Department of Electronics and Intelligent Systems Engineering, School of Industrial Engineering, Tokai University
9-1-1, Toroku, Kumamoto 862-8652, Japan

Abstract- **We propose a method for detecting vehicle positions and their movements by using thermal images obtained through an infrared thermography camera. Our method specifies the area of moving vehicles based on the variations of pixel values, i.e. the standard deviations of pixel values along the time direction of spatio-temporal images. It also specifies vehicle positions by applying the pattern recognition algorithm which uses Haar-like features per frame of the images. The results of our experiments show that the information about both vehicle positions and their movements can be obtained by combining those two kinds of detection. Our method has the advantage of distinguishing between smooth traffic flows and heavy ones. In addition, there is a possibility that traffic accidents, vehicle troubles, and illegal parking can be detected with our method. We also expect to realize a robust automatic monitoring for road traffic around the clock even in changeable weather with our method of using thermal images.**

I. INTRODUCTION

It is a pressing matter how to develop the vision-based traffic monitoring systems in the field of ITS (Intelligent Transportation Systems). They have the advantage of measuring what could not be measured by conventional vehicle detectors: vehicle positions, vehicle lengths, and moving vehicle on multilane. Therefore, by use of the vision-based traffic monitoring systems, we can make up both a new traffic control strategy designed to reduce traffic jams and an automatic traffic monitoring system designed to find traffic incidents without time delay.

The present research level of the vision-based traffic monitoring systems is high enough to detect vehicles robustly around the clock [1][2]. However, there are some defects as follows.

Many of the conventional methods detect the bodies of vehicles only in the daytime, and nighttime detection is adopted in much fewer cases. Yoneyama et al. [2] pointed out that most of the daytime detection methods lose their accuracy when they are directly applied to nighttime detection. Therefore, generally adopted methods have been those of detecting the headlights or the taillights of vehicles at nighttime and of preparing two algorithms separately for daytime and nighttime detection [1][2]. It is generally difficult to measure the vehicle sizes at nighttime. In the conventional method of detecting taillights [2], the vehicle lengths can be measured only at the limited camera angles because the vehicle length is estimated by the triangle of a pair of taillights and the vehicle front edge detected as the gray level difference between the head of the vehicle and the headlight reflection area on the road. What is the most important for both traffic signal control and traffic simulation is to be able to measure the vehicle sizes and to specify the number of large-sized vehicles in the detected area around the clock. It is especially true because the saturation flow rate used as one of the fundamental values for them is largely influenced by the rate of large-sized vehicle mixing.

Vehicle cast shadows at daytime impair the vehicle detection of the adjoining lane. Therefore, a method of eliminating vehicle cast shadows has already been proposed [3]. However, it has the disadvantage of restricting the elimination of the vehicle cast shadows to limited camera angles.

In the conventional methods of vehicle detection with visible light cameras, it is difficult to detect vehicles with high accuracy in bad weathers such as fog, snow, and heavy rain. However, traffic accidents and traffic jams are most likely to happen under such circumstances. Therefore, it is a pressing matter for us to develop a method designed to detect vehicles with high accuracy under all circumstances.

In this paper, we propose a method for detecting vehicle positions and their movements by using thermal images obtained through an infrared thermography camera instead of a visible light camera. The thermal images of vehicles are expected to detect the shapes of vehicles robustly regardless of any environment around the clock.

First, we refer to our observations in Section II which show how to obtain clear visions of vehicles even in snowy and deep foggy weather. Second, the algorithm developed in our method is explained in Section III. Third, our experiments at daytime with vehicle cast shadows are explained. Then we refer to the experimental results which show that our method offers reliable information about both vehicle positions and their movements without the influence of the vehicle cast shadows. Therefore, this information is useful for an

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_40,

automatic monitoring for road traffic. Finally, we have conclusions in Section V.

II. Advantages of Thermal Images in the Vehicle Detection for Bad Environments

The infrared thermography camera used in our detection is TVS-200 [4]. The frames of the infrared thermal images are transmitted to a notebook personal computer with the 1/60 seconds interval through the IEEE1394 interface. The infrared thermography camera and the notebook personal computer are shown in Fig. 1.

We have confirmed through our observations that the obtained thermal images are contrasted highly enough to detect the shapes of vehicles even in bad weather.

We have obtained both thermal images and visible light images in snowy and deep foggy weather on a mountain road of Aso in Kumamoto, Japan. Images of moving vehicles taken from the infrared thermography camera are shown in Fig. 2 (a) (b) and its image taken from a visible light camera is shown in Fig. 2 (c). The outlines of the vehicles are clearly seen in Fig. 2 (a) (b) while only the position of the fog lamps is seen in Fig. 2 (c). The bars of the right side in Fig. 2 (a) (b) show the temperature scale.

The infrared thermography camera, therefore, can be effectively used for vehicle detection under bad circumstances like snowy and deep foggy weather.

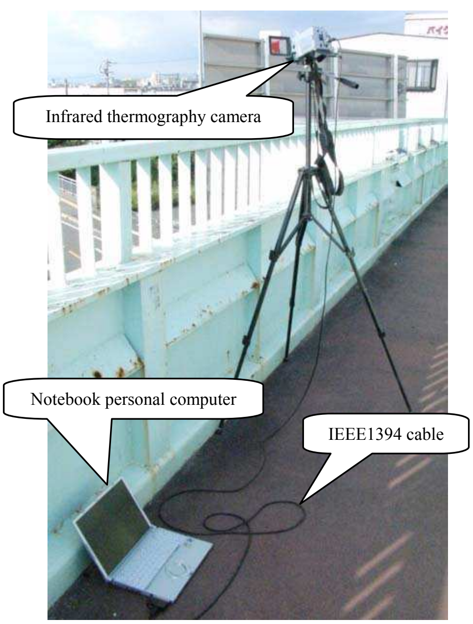


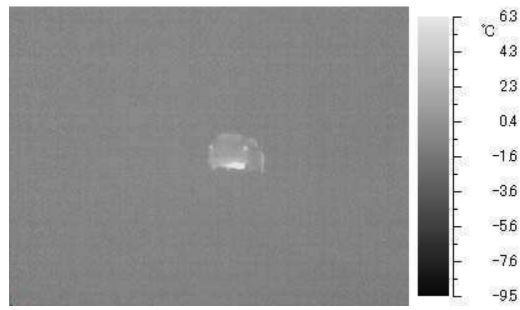
Fig. 1. The infrared thermography camera and the notebook personal computer on a pedestrian bridge.

III. A Method for Detecting Vehicle Positions and Their Movements

The proposed method is applied to the detection of vehicles from the traffic thermal images obtained through the infrared thermography camera set up at the height of a pedestrian bridge.

A. Spatio-temporal image processing

The standard deviations of pixel values for n frames in the past are calculated from the thermal images as follows.



(a) Thermal image (back side view)

(b) Thermal image (front side view)

(c) Visible light image (front side view)

Fig. 2. Thermal images and a visible light image in snowy and deep foggy weather.

$$\sigma(x,y,t_c)=\sqrt{\frac{\sum_{t=t_p}^{t_c}(f(x,y,t)-\mu(x,y,t_c))^2}{n}}, \quad (1)$$

where $\sigma(x, y, t_c)$ is the standard deviation at coordinates (x, y) for n frames in the past from time t_c, $f(x, y, t)$ is the pixel value at coordinates (x, y) at time t, and $\mu(x, y, t_c)$ is the mean value at coordinates (x, y) for n frames between time t_p and time t_c.

The standard deviations indicate the variations of pixel values along the time direction. By computing the standard deviations of all pixels in the frame, we can distinguish between the area of moving vehicles and that of the background or stopped vehicles based on n frames in the past. In other words, if the standard deviation is not zero, it means that the pixel value is changed by a moving vehicle. Conversely, if the standard deviation is zero, it means that the pixel value maintains the same value in the area of the background or stopped vehicles. When the standard deviation is more than or equal to s_t, the pixel is actually assumed to be in the area of moving vehicles because each pixel value includes a noise.

B. Vehicle pattern recognition using Haar-like features

The pattern recognition algorithm using Haar-like features was proposed by P. Viola et al. [5]. P. Viola et al. proved in their paper[5] the effectiveness of the algorithm concerning the experiments of face detection. By changing the object of pattern recognition, this algorithm can be applied to the detection of other objects like vehicles. To detect vehicles in images, we have used two types of images: the positive samples including a vehicle and the negative samples including no vehicle. Then, the machine learning has been done by using them, and a multistage cascade of classifiers has been constructed. We can make a pattern recognition by using the obtained multistage cascade of classifiers.

The images we have collected contain the windshield and its surroundings which clearly show the front view of a vehicle as positive samples while they contain no vehicle as negative samples. By using the upper part of the vehicle such as the windshield and its surroundings as the target of pattern recognition, we can make a robust detection of vehicles even when they are stopped one after another with a short distance. We have conducted training with these sample images to obtain a multistage cascade of classifiers, and finally managed to make the pattern recognition of vehicles with the obtained multistage cascade of classifiers. In our vehicle pattern recognition, we have used the extended algorithm proposed by R. Lienhart et al. [6] for the paper [5].

C. Combination of two kinds of information: vehicle positions and their movements

By combining the two kinds of processing in the same frame of images as described in sections III-*A* and III-*B*, the position of each vehicle can be specified, and its movement can be classified, too.

Each vehicle speed can be classified based on the ratio of the area of the moving vehicle in the rectangle which shows the outline of detected vehicle. In our method, we have classified the three categories: If the ratio is less than a_1%, the vehicle is assumed to be stopped. If the ratio is more than or equal to a_1% and less than a_2%, the vehicle is assumed to be running at low speed. If the ratio is more than or equal to a_2%, the vehicle is assumed to be running at high speed.

IV. EXPERIMENTAL RESULTS

We have developed our algorithm with Visual C++ 2008 and the computer vision library OpenCV [7].

Fig. 3 shows one frame of thermal images. The frame size of collected images is 320x240 pixels. Fig. 4 shows some examples of the positive sample images, and Fig. 5 shows some examples of the negative sample images. The number of positive sample images and negative ones used in our experiments are 10,106 and 4,000 respectively. Each positive sample image is resized to 12x8 pixels for the training. The number of the stages of the classifiers obtained through the training is 14. We have assumed in the experiments that n is 30, s_t is 3.0, a_1 is 10.0, and a_2 is 40.0.

Fig. 6 shows an example of spatio-temporal images with the inscription of space-axes xy and time-axis t. Fig. 7 shows the results of the detection. The five images of Fig. 7 (a-1), (a-2), (a-3), (a-4), and (a-5) show the interim results we have got by combining the two kinds of processing, i.e. spatio-

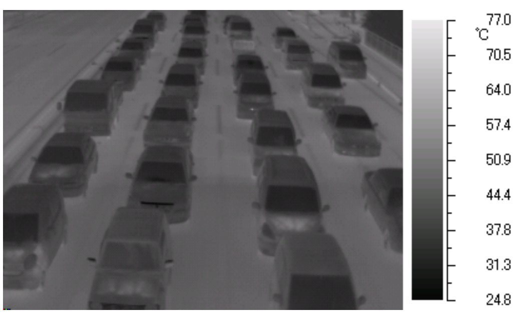


Fig. 3. A frame of thermal images.

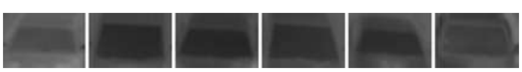


Fig. 4. Examples of positive sample images.

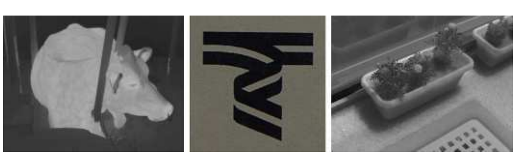


Fig. 5. Examples of negative sample images.

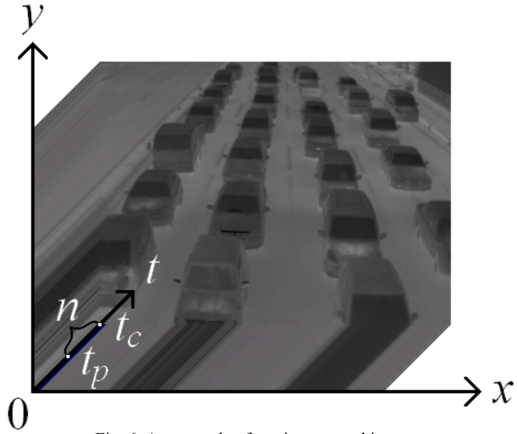


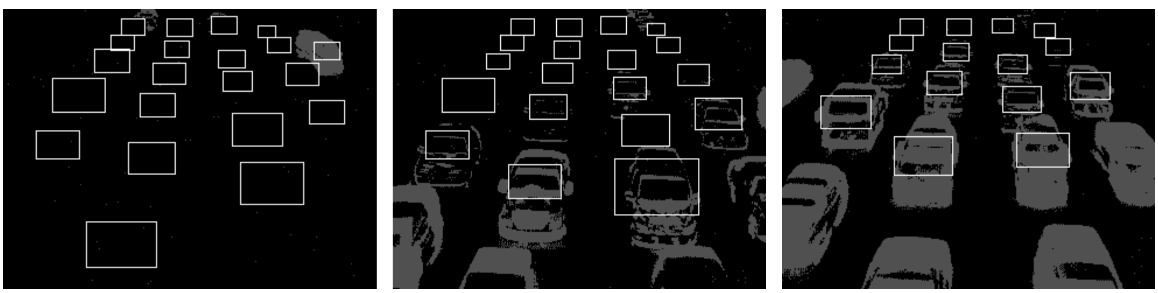
Fig. 6. An example of spatio-temporal images.

temporal image processing and vehicle pattern recognition. The areas of gray pixels show those of moving vehicles specified by the processing described in the section III-*A*, and the white rectangles show the outlines of vehicles specified by the pattern recognition described in the section III-*B*. The five images of Fig. 7 (b-1), (b-2), (b-3), (b-4), and (b-5) show the final results of the detection which have enabled us to specify each vehicle position and its movement. The five images of Fig. 7 (a-1), (a-2), (a-3), (a-4), and (a-5) correspond to those of Fig. 7 (b-1), (b-2), (b-3), (b-4), and (b-5), respectively. The dotted lines, the thin lines, and the bold lines in the final results in Fig. 7 show three categories of vehicles: stopped ones, slowly moving ones, and fast moving ones, respectively.

In the interim results of Fig. 7, there is no connected components of gray pixels for stopped vehicles because the standard deviations of pixel values are distributed less than s_t. When a vehicle begins to move and runs at low speed, the standard deviations of pixels, which are more than or equal to s_t, begin to appear in the images. Consequently, a part of the rectangle which shows the result of vehicle pattern recognition contains the connected components of gray pixels. In addition, when the vehicle runs at high speed, the area of gray pixels spread more widely, and most of the rectangle area becomes gray pixels.

By combining the results of spatio-temporal image processing and the pattern recognition, we can classify the speed of each vehicle. In our experiments, we classify vehicle speed into three categories as described above.

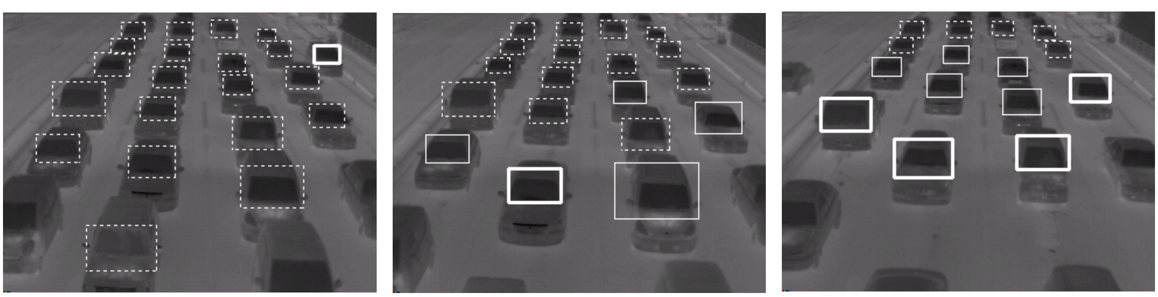
Fig. 7 (b-1) shows the time soon after the green light is turned on. All vehicles except the one on the left-turn lane are stopped. Fig. 7 (b-2), (b-3), and (b-4) show the passage of time after the green light is turned on. Vehicles begin to move and increase their speed gradually from the head to the back of the line. Fig. 7 (b-5) shows the line of vehicles on the right-turn lane is made longer. As shown in Fig. 7, each



(a-1) A result of both spatio-temporal processing and pattern recognition

(a-2) A result of both spatio-temporal processing and pattern recognition

(a-3) A result of both spatio-temporal processing and pattern recognition



(b-1) A detection result for vehicle positions and their movements

(b-2) A detection result for vehicle positions and their movements

(b-3) A detection result for vehicle positions and their movements

Fig. 7. Thermal Image detection results.

vehicle position and its movement can be detected in real time with our method.

Fig. 8 shows a visible light image taken at the same time. The vehicle cast shadows extend to the adjoining lane in Fig. 8 while there is no influence of vehicle cast shadows on the detection of vehicles in Fig. 7.

V. Conclusions

First, we have confirmed through our observations that the thermal images obtained through an infrared thermography camera offer the images of vehicles clear enough to detect their shapes even under bad circumstances like snowy and deep foggy weather.

Then, we have proposed a method for detecting vehicle positions and their movements by using thermal images. Our method specifies the area of moving vehicles based on the variations of pixel values, i.e. the standard deviations of pixel values along the time direction of spatio-temporal images. It also specifies vehicle positions by applying the pattern recognition algorithm which uses Haar-like features per frame of the images. The results of our experiments show that the information about both vehicle positions and their movements can be obtained by combining those two kinds of detection. Our method has the advantage of distinguishing between smooth traffic flows and heavy ones. In addition, there is a possibility that traffic accidents, vehicle troubles, and illegal parking can be detected with our method. We also expect to realize a robust automatic monitoring for road traffic around the clock even in changeable weather with our method of using thermal images.

Acknowledgment

We wish to express our gratitude to Prof. Shuhei Takada of Tokai University, a researcher of American culture & literature, for having carefully corrected this paper.

References

[1] R. Cucchiara, M. Piccardi, and P. Mello, "Image Analysis and Rule-Based Reasoning for a Traffic Monitoring System," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 1, No. 2, pp.119-130, June 2000.

[2] A. Yoneyama, C. H. Yeh, and C. -C. J. Kuo: "Robust Vehicle and Traffic Information Extraction for Highway Surveillance," *EURASIP Journal on Applied Signal Processing*, Vol. 2005, Issue 14, pp. 2305-2321, 2005.

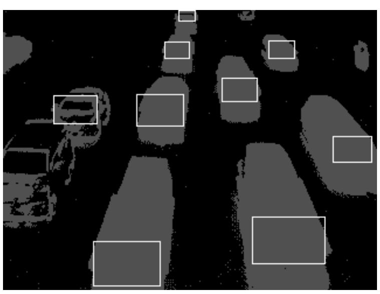
[3] A. Yoneyama, C. H. Yeh, and C. -C. J. Kuo, "Moving Cast Shadow Elimination for Robust Vehicle Extraction based on 2D Joint Vehicle/Shadow Models," *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 229-236, July 2003.

[4] http://www.avio.co.jp/english/products/tvs/pdf/tvs200_4p_e_0520.pdf

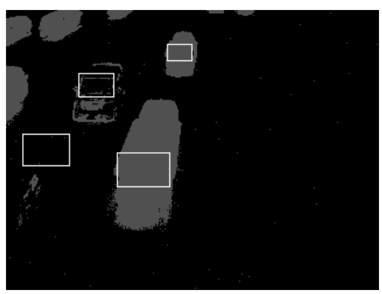
[5] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp.511-518, December 2001.

[6] R. Lienhart and J. Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection," *Proceedings of the IEEE 2002 International Conference on Image Processing*, Vol. 1, pp.900-903, September 2002.

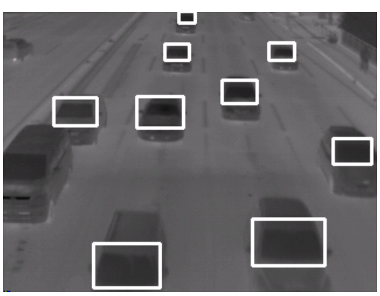
[7] http://sourceforge.net/projects/opencvlibrary/



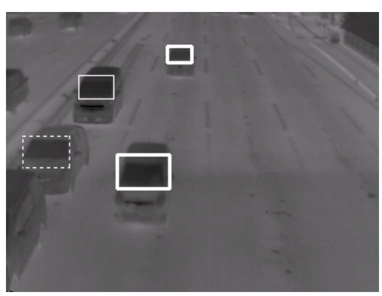
(a-4) A result of both spatio-temporal processing and pattern recognition



(a-5) A result of both spatio-temporal processing and pattern recognition



(b-4) A detection result for vehicle positions and their movements



(b-5) A detection result for vehicle positions and their movements

Fig. 7. (Continued).

Fig. 8. A frame of visible light images.

Virtual Oxygen Sensor Implementation Using Artificial Neural Networks

Thiago Richter[1], Amaury Ferrari Oliveira[2] and Ivan Nunes da Silva[3]

[1] *Delphi Automotive Systems, Piracicaba, Brazil, thiago.richter@delphi.com*
[2] *Delphi Automotive Systems, Piracicaba, Brazil, amaury.oliveira@delphi.com*
[3] *Department of Electrical Engineering, University of São Paulo, São Carlos, Brazil, insilva@sel.eesc.usp.br*

Abstract - ***The Engine Management System is the system responsible for controlling the combustion in an automotive engine that is a very complex system with various sensors and actuators involved. This paper presents the use of artificial neural networks to estimate values of the oxygen sensor, which is one of the main sensors used by the Engine Management System and is directly related to the control of combustion and emissions. Together with the costs reduction, the control of emissions has received great focus by the automotive industry, where one of the reasons is actually the environmental issue.***

I. INTRODUCTION

The automotive industry is one of the most important sectors in Brazilian's economy and in the world, with net revenue of more than $ 52 billion only in Brazilian vehicles production in 2007. Considering only the automakers participation the Brazilian automotive industry corresponds to 18% of the industrial GDP [1].

In recent years the automotive industry has been forced to improve the performance of their produced vehicles and reduce their costs. One of the landmarks of this transformation was the development of the oxygen sensor. With the oxygen sensor was possible to develop the Engine Management Systems which enables, among its various functions, controlling the combustion and consequently vehicle's emissions.

The development of this work was motivated by the constant research by automotive industry for costs reduction and technological advances, as well as the increasing progress in the use of intelligent systems, including the development of virtual sensors for various purposes [2,3,4]. This study proposes the use of intelligent systems architectures for virtual oxygen sensing of vehicles and can therefore replace a physical sensor for a virtual sensor, estimating the reading of the oxygen sensor, creating a virtual oxygen sensor using supervised artificial neural network that can be inserted into the engine management system after trained.

In this paper, we describe concepts of artificial neural networks, engine management system and oxygen sensors. Soon after, it is showed the intelligent system developed. Finally, the results and conclusion about the proposed work are outlined in the last section.

II. ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANN) are computational models inspired by the human brain and have the ability to acquisition and maintenance the information. The ANN can also be defined as a set of processing units (neurons) that are interconnected by a large number of synapses [5,6]. Fig. 1 gives the general model of McCulloch's artificial neuron, where x are the inputs of ANN, w are their weights, θ is the threshold value, f is the activation function of the neuron and y is the output.

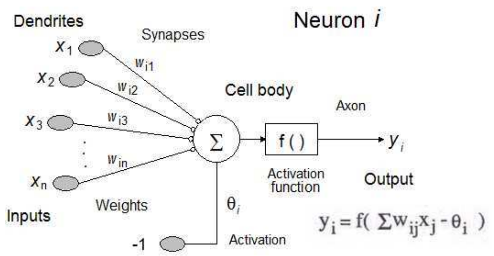


Fig 1. General model of artificial neuron.

The ANN has as its main features the ability to adapt and learn, including learning through examples, ability to generalize, group or organize data, self-organizing, tolerate failures and not require the mathematical modeling of the process. Several are their areas of application [5]: patterns recognition, functions approximation, forecasting/estimating, system optimization, clustering, associative memories and control systems.

Several topologies are used in ANN, among them the multilayer perceptron (MLP) topology stands out. The MLP has feedforward architecture and consists of an input layer (x_1, x_2,...,x_n), one or more hidden neural layers (HL), a neural output layer (OL) and output (y) as shown in Fig. 2. One of the main applications of the MLP is the function universal approximations and pattern recognition. The universal approximation, based on artificial neural networks, is able to perform non-linear mapping [7], listing the inputs to the outputs.

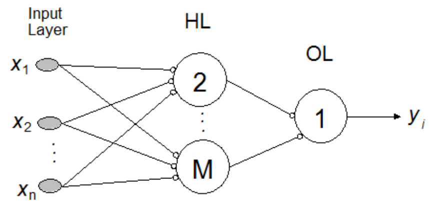


Fig. 2. An MLP architecture model.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_41, © Springer Science+Business Media B.V. 2010

III. Engine Management System

Automotive control systems are dynamic systems involving the interaction of continuous and discrete components (Fig. 3) and are called hybrid systems [8]. The Engine Management System (EMS) is a typical example of hybrid system, since it is a real-time system [8].

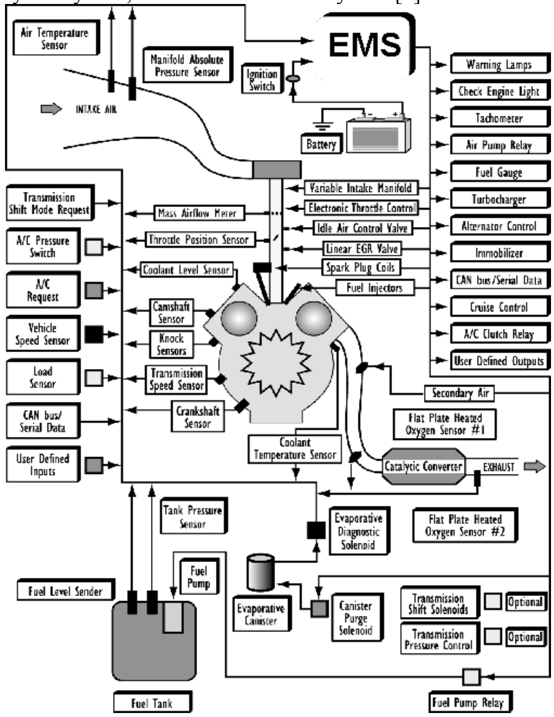


Fig. 3. Example of inputs and outputs of an EMS.

An EMS controls the internal combustions of the engine, which represents a highly complex problem of controlling [2], where the engine contains many elements that are characterized by intense integration of nonlinear functions. They can have as input various sensors, such as the oxygen sensors, temperature sensors, pressure, altitude and diagnostic sensors, among others, using this information to control the ignition system.

The use of EMS allows better control, better sensitivity, greater fuel efficiency and lower levels of pollution. Modern systems can control other automotive components or systems, such as the turbo or exhaustion, and communicate with other systems, such as the transmission system, cooling, etc [9].

Due to the complexity of the EMS, one of the biggest challenges in its implementation is the long time needed to perform the calibration of the system and it is extremely important to have knowledge of an expert.

IV. Oxygen Sensor

The Air-Fuel ratio (A/F ratio) refers to the proportion of air and fuel used during combustion and it is often defined as being an air excess factor. This proportion is called ideal stoichiometric ratio [9]. Already factor Lambda, is defined as a unity coefficient that corresponds to a rate of 14.7 liters of air per 1 liter of fuel that is 14.7:1 A/F, to gasoline engines in normal temperature and pressure [10]. These sensors are also called lambda sensors in Brazil.

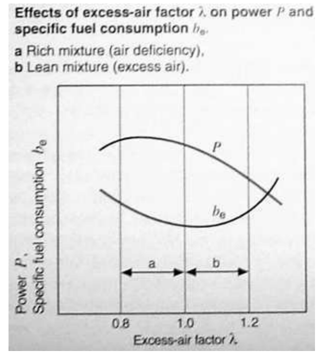


Fig. 4. Lambda value – Power *vs* Fuel Consumption [9].

A larger amount of gasoline results in lambda less than one, called rich mixture, while a larger portion of air results in a lambda greater than one, called weak or poor mixture (Fig. 4). The greatest power is produced when lambda is approximately 0.9 and the minimum fuel consumption occurs when lambda is close to 1.1; for more torque with less consumption the expected result for lambda is 1.0, corresponding to the rate of 14.7:1, and the acceptable error for oxygen's sensor is 0.2% [10, 11].

According to [11], many factors are responsible for a non-optimal combustion and result in reduced emissions of harmful gases but we can highlight the A/F ratio as one of the main reasons. To obtain the most efficient conversion of CO, HC and NOx gases (Fig. 5) in a three-way catalytic converter the ideal A/F mix must have lambda equal to 1.

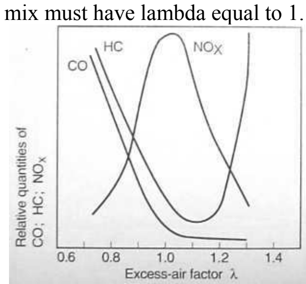


Fig. 5. Gases emissions [9].

Currently, it has been large the numbers of environmental laws that limit the emissions that the auto industry must achieve to serve the markets, some examples are [12]:

• Legislation: PROCONVE (Brazil), Euro IV and Euro V, U.S. Environmental Protection Agency (EPA).

• Standards: Tier 2 (EPA), LEV (Low Emission Vehicle), ULEV II (Ultra Low Emission Vehicle) and SULEV (Super Ultra Low Emission Vehicle).

In Brazil the resolution 18/86 regulates the vehicles emissions [13].

V. Development

Since the raw data is quantitative and made available on large batches, the application of RNA is indicated. These data come from readings of available variables on the EMS, collected from an 1.8cm^3 engine size vehicle within 8 valves, at dynamometers from Delphi Brazil Technical Center, without special arrangements on the engine to make it functional, and using the final test conducted by the Delphi's engine calibration Brazilian team. The engine will be calibrated without fixed variables, trying to reproduce ordinary driver schedule at dynamometer level. This test will be measuring 42 variables.

The methodology used to train the system was the cross validation technique, using 75% of available data to make the training, where 75% of this data was used to train and 25% to validate, and the other 25% to check if the network is converging satisfactorily after training. All topologies have been trained with the same data sets, and all tests used the same data.

In order to simulate the delayed feedback system (EMS), where the current value of lambda is a consequence of all the other variables, the time of lambda when related to other variables has been moved to t+1.

In Fig. 5, 6 and 7, graphs of collected variables are presented, which were used on training. Each variable was normalized between 0 and 1 using their minimum and maximum values. Fig. 5 shows an enlarged graph only with the variable containing the lambda value, which is the target value for the neural network output developed. The typical behavior of the oxygen sensor on a calibrated vehicle could be observed with the zoom, where lambda values maintain switching between rich and poor mixture controlled by EMS that is constantly trying to balance the A/F ratio.

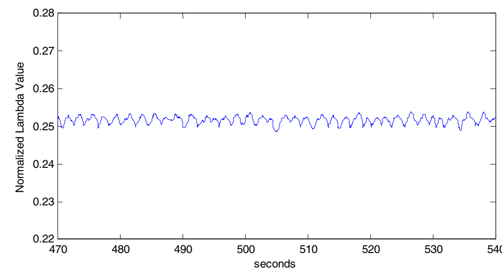


Fig. 5. Output target detail.

Fig. 6 shows the whole universe of available output targets, without zoom in as in Fig. 5.

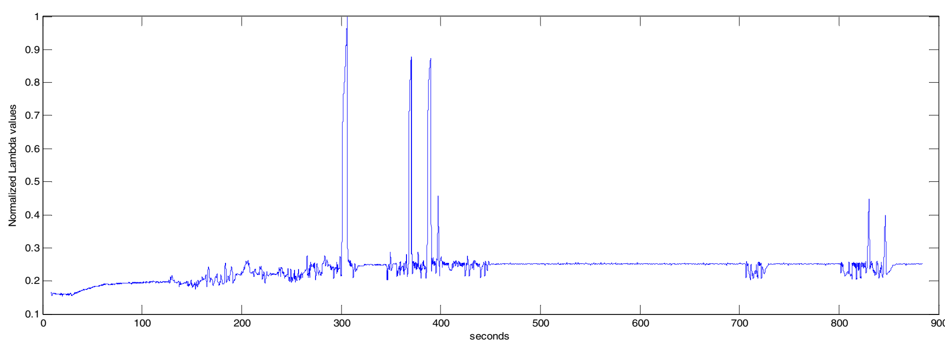


Fig. 6. Output target from data map 1.

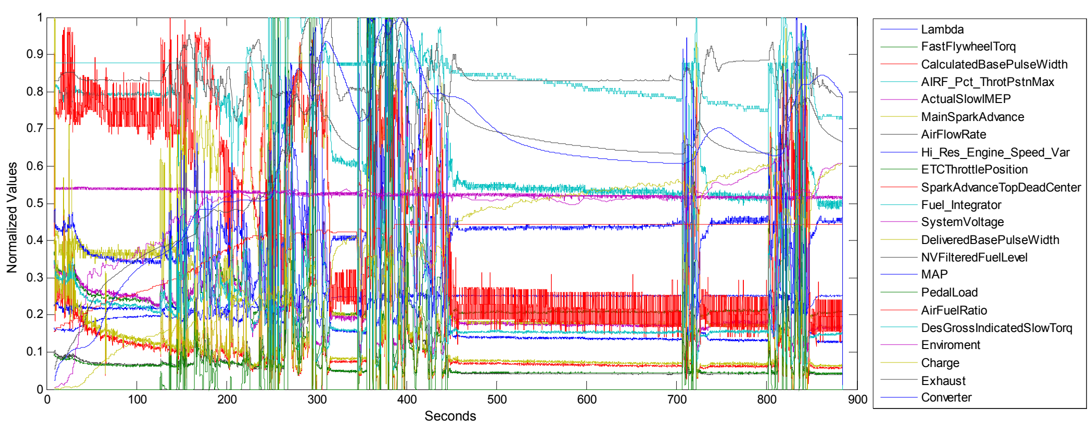


Fig. 7. Graph with some normalized variables.

As data maps are extensive, more than 85 thousand lines, there was a need to discrete this data to make the training and testing. Two approaches were used. Within the first approach, the entire available universe of data was discretized on a percentage from the whole universe. The second discretization was performed using subsets defined by the user and also considering a percentage of the data set.

The variables used at training and testing were divided into three sets. The first set used all available variables, the second set used the less linear variables (Fig. 7), observing the graphs from Fig. 7. The third set was based on the expert knowledge from calibration's engineers team.

The developed system can create a loop of training and testing, using variables that allow selecting the network inputs. Some allowed settings are: definition of data maps that should be used, type and percentage of discretization, the input variables selection, the goal of the network, the size of hidden neural layer, training type, maximum number of epochs to train, activation and training functions.

VI. Results

The analysis took into account, primarily the performance value of the tests after the neural networks have been trained and then, to prevent the effects of overfitting and underfitting, each graphics were seeing and allowing evaluation of which are the best fit models. Altogether more than 2,300 neural networks were generated, ranging up their inputs and features, as described in Section 5.

Below follows analyzes of both types of discrete data used at training. First trainings and tests were performed using entire universe of data and then the training and tests involved the second type, using discrete data from subsets. The best results were achieved using the following variables shown in Table 1.

TABLE I
Best Results Variables

1	Oxygen Sensor	26	System Voltage
2	Torque	27	Delivered base pulse width
6	Pulse width	28	Fuel level
7	Throttle-body air flow	30	MAP
8	IMEP	31	Pedal load
9	Main Spark Advance	32	Air fuel ratio
10	Air flow rate	34	Slow torque indication
12	Engine speedy	37	Environment temp.
13	Throttle-body position	40	Charge
14	Spark Advance	41	Exhaust temp.
15	Fuel Integrator	42	Converter temp.

Table 2 shows the Top two performances for each type of discretization (data range) and data map as well as their mean relative errors and some main characteristics, where "Vars" means variables, "Train" means training method and "HN" means the neurons number in the hidden layer.

TABLE II
Best Performances of Each Map and Type of Discretization

# Top.	Data range	ANN main characteristics	Rel. Error
1.320	Map 1 Whole Univ.	Data range: [9 888]; Vars: 42; Train.: trainbr; HN: 10	1,9801%
1.309	Map 1 Whole Univ.	Data range: [9 888]; Vars: 42; Train.: trainbr; HN: 10	2,0904%
1.706	Map 2 Whole Univ.	Data range: [9 888]; Vars: 42; Train.: trainlm; HN: 15	0,3400%
1.718	Map 2 Whole Univ.	Data range: [9 888]; Vars: 42; Train.: trainlm; HN: 20	0,3439%
2.194	Map 1 Subset	Data range: [446 534]; Vars: 22; Train.: trainlm; HN: 20	0,2062%
2.198	Map 1 Subset	Data range: [446 534]; Vars: 22; Train.: trainlm; HN: 13	0,2063%
2.455	Map 2 Subset	Data range: [271 359]; Vars: 22; Train.: trainlm; HN: 25	0,1518%
3.336	Map 2 Subset	Data range: [271 359]; Vars: 22; Train.: trainlm; HN: 55	0,1583%

A. *Training Using Whole Data Universe*

Whereas the test uses the entire universe of available data, we can conclude that the result shown in Fig. 8 was satisfactory for the data map 1, getting a mean relative error ($\bar{e}_r$) of 2.0904%.

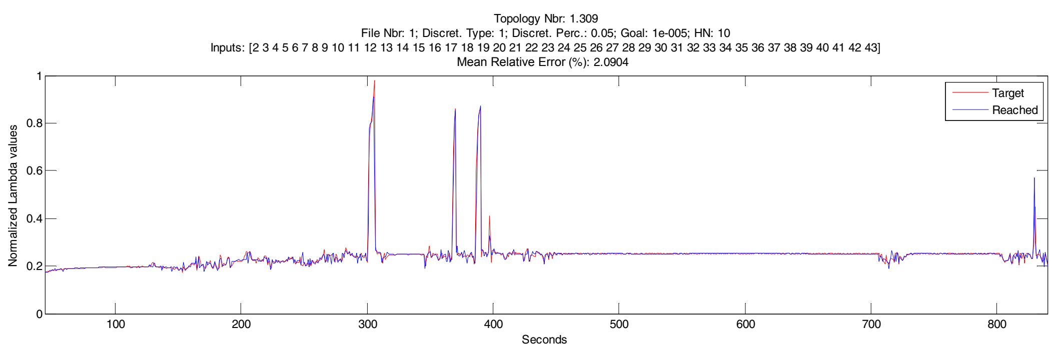


Fig. 8. Validated result example using data map 1 and whole universe.

Fig. 9 shows an expansion of other results with the same map as in Fig. 8, allowing detailed view of the result. Fig. 10 shows the absolute errors, where the absolute errors values are the difference between the target value and achieved.

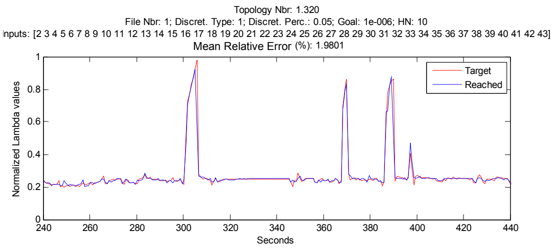


Fig. 9. Result details of the graph shown at Fig. 8.

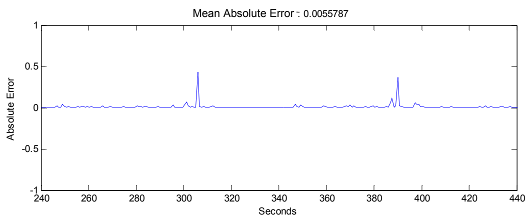


Fig. 10. Absolute errors of the Fig. 9.

Fig. 11 shows the enlarged test result which was used the whole universe with data map 2, where the $\overline{e}_r$ was 0.3439%, and Fig. 12 presents the relative errors.

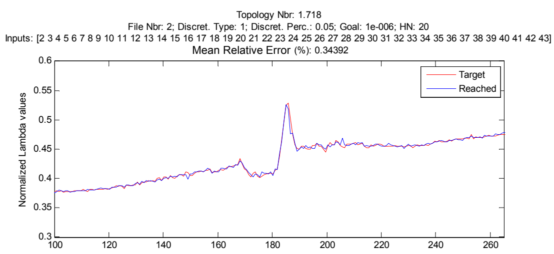


Fig. 11. Result details using data map 2 and whole universe.

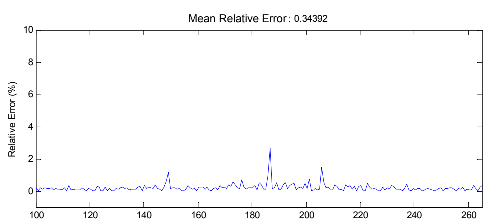


Fig. 12. Relative errors of Fig. 11.

B. Training Using Subsets of Data Universe.

At Fig. 13, a satisfactory result is shown, with relative errors below 0.45%, where the real value of lambda is irregular and the system succeeded at bringing the lambda value using data map 1.

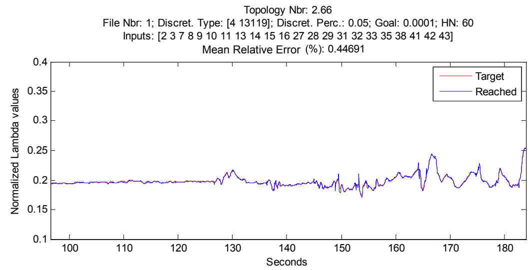


Fig. 13. Result example using data map 1 and subsets.

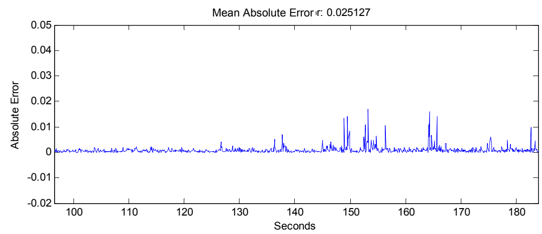


Fig. 14. Absolute errors of the Fig. 13.

The following example (Fig. 15) illustrates the typical oxygen sensor behavior, noting that the behavior was maintained and with good accuracy, with a mean relative error of 0.21%.

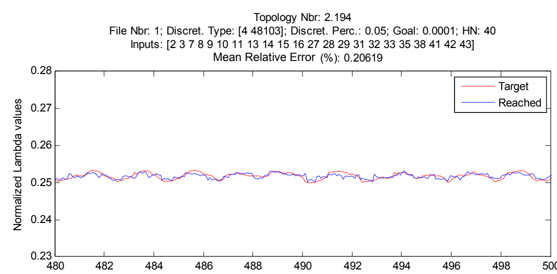


Fig. 15. Results detail using map 1 and subsets.

At Fig. 16 and 17 may be observed, respectively, absolute errors and relative errors concerning to Fig 15.

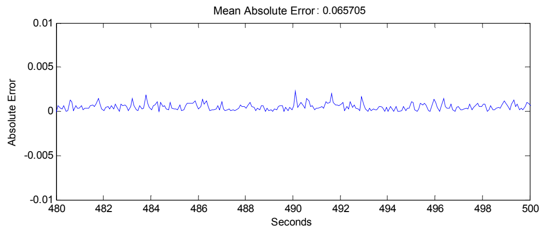


Fig. 16. Absolute errors regarding Fig. 15.

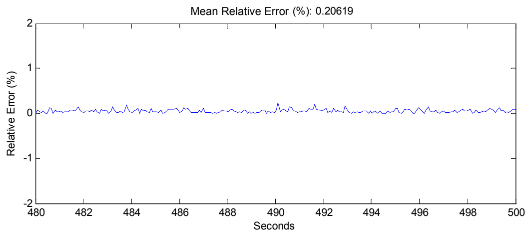


Fig. 17. Relative errors of the Fig. 15.

VII. Conclusions

The virtual oxygen sensor developed using MLP neural networks reaches a closer to real results given the complexity of an engine system, reaching low mean relative errors in hundreds of topologies.

Neural networks topologies using the whole universe of data discretized can be used, although using subsets of the universe to train achieves better results. Using the whole universe to train, almost 350 topologies out of 792 reached $\bar{e}_r$ less than 1%. Already at subsets training, more than 600 topologies of both maps have achieved $\bar{e}_r$ less than 0.5%.

As a suggestion, for future work, applying the trained neural network on a car level engine management system, furthermore, it is necessary to decrease the error evaluating other topologies, the best variables and excluding which are the noise variables.

Acknowledgment

The authors would like to thank everyone who cooperated with this work, especially Elisabete M. D. Oliveira and some employees of Delphi Automotive Systems (Fabio A. M. Fantinato, Orlando Volpato Filho and Sandro Souza).

References

[1] ANFAVEA, 2008, “Brazilian auto industry yearbook – 2008,” in portuguese, Brazilian automakers association, São Paulo, 2008.

[2] R. J. Howlett, “Monitoring and control of an internal combustion engine air-fuel Ratio using Neural and Fuzzy Techniques,” Department of Electrical and Electronic Engineering, University of Brighton, UK, 1998.

[3] P. Amato, N. Cesario, M. Di Meglio and F. Pirozzi, “Realization of a virtual lambda sensor on a fixed precision system,” Proceedings of the Design, Automation and Test in Europe, IEEE, vol. 3, p. 192-197, 2005.

[4] S. Kamat, V. Diwanji, J. Smith, H. Javaherian and K. P. Madhavan, “Virtual sensing of SI engines using recurrent neural networks,” 2006 SAE World Congress, Electronic Engine Controls, SAE, 2006

[5] S. Haykin, “Neural networks,” Prentice-Hall, 1999.

[6] Z. L. Kovács, “Artificial neural networks: concepts and applications,” in portuguese, Edição Acadêmica, São Paulo, 1996.

[7] R. M. Golden, “Mathematical methods for neural network analysis and design,” MIT Press, 1996.

[8] DELPHI, “A model-based environment for production engine management system (EMS) development,” SAE, 2001.

[9] BOSCH, “Automotive technology handbook,” in portuguese, Edgard Blucher, São Paulo, 2005.

[10] DELPHI., “Planar oxygen sensor: application manual.,” Delphi Automotive Systems, Piracicaba, 2005.

[11] BOSCH, “Gasoline-engine management,” SAE, 1999.

[12] DELPHI, “Delphi high function engine management system,” Delphi Automotive Systems, Troy, 2006.

[13] BRASIL, Resolution number 18, May 6th, 1986. “Air pollution control program for auto vehicles – PROCONVE,” Union Official Daily, Brasília, 1986.

Pattern-Based Usability Analysis: Comparison with a Case-Study

Dmitry Zub, Erki Eessaar
Tallinn University of Technology, Estonia
dmitri.zub@mail.ee, eessaar@staff.ttu.ee

Abstract **- In our previous paper, we presented a method for evaluating the usability of e-learning systems. Also a general method was proposed that could be used to evaluate different aspects of software systems using different kinds of patterns. In this paper, we describe the results of a research, during which we looked through some evaluations of e-learning systems. We did it in order to find a published research (control study) with which we could compare our method of analysis. The paper describes the study we have selected for our re-evaluation. We also present the results of our study and compare these with the results of the control study.**

I. INTRODUCTION

There are a lot of requirements that we could present to a software system. Along with functional requirements there exist non-functional requirements. One desired non-functional property of systems is *usability*. If usability of a system is good, then it means that a user:

- can learn to use the system without any problems,
- can quickly complete frequent tasks,
- can easily recall how to use the system even if he/she has not used that for some period of time,
- can quickly recover from the errors that occur during interacting with the system.

Usability is significant for different types of software systems, but there are some areas where it is especially important. One such area is *educational software*. Students should not wonder how the software works because it distracts them from studying. In our previous paper [1], we suggested a method for analyzing the usability of several e-learning systems. The method proposes to evaluate systems by using the Analytical Hierarchy Process (AHP). The evaluation criteria are different usability and e-learning patterns. The evaluators have to determine how well the systems conform to these patterns. *The first goal* of this paper is to analyze the usability of some e-learning systems by using our proposed method and publish the results of this study. *The second goal* is to compare our results with the results of the already published analysis of these systems. This way we can see whether our method is effective and compare it with another method.

The organization of this paper is the following. In the section II, we will say a few words about some evaluation studies we have found in the literature. In the next section, we will describe the research (control study) that was selected for our analysis and with which we will compare the results of our method. In the section IV, we will describe the steps of our method. We are going to use this method in order to evaluate the usability of a set of e-learning systems and compare the results with an existing study. In the next section, we describe the steps that we have done in our study. In the section VI, we will present the results of our study. In the next section, we will shortly describe the results of the control study. In the section VIII, we will compare the results of our study and the control study and discuss them. Finally, we will describe our future work and draw conclusions.

II. RELATED WORK

In this section, we are going to give an overview of some existing researches about e-learning and usability.

Bafotsou and Mentzas [2] analyzed web-based collaborative systems (44 products and research prototypes in total). They provided a list of basic services that are included in the systems like e-mail, whiteboard, real- and not real-time discussions etc. Comparison of functions of the systems was made. They also divided the systems into five functional categories.

Karoulis [3] evaluated the usability of LEGO educational programming environment called "RoboLab". Cognitive walkthrough and heuristic evaluation methods were used in the analysis.

Colace, De Santo and Vento [4] evaluated 15 on-line learning platforms in relation to several groups of parameters like systems requisites, training resources and course management, user management, and services offered to users.

Ngai and Chan [5] used AHP to evaluate three knowledge management tools. They did not use patterns as criteria.

Kljun et al. [6] compared the papers that evaluate different learning management systems. Out of 31 analyzed papers, 17 compared features of the systems.

Komarkova et al. [7] evaluated the usability of GeoWeb Sites. However, this system is not an e-learning system.

These papers were not suitable for our study because of the following reasons. Our method is based on the use of patterns in order to compare different systems. On the other hand, most of these papers compared the *features* of some set of systems and it is hard to perform this kind of analysis with our method. The reason is that each pattern describes "a solution to problem" and there are no patterns that specify the functions that a good system should have. In addition, some of the studies considered more systems than we can include in our analysis. We must have access to all the systems that are

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_42,

evaluated in the previous study in order to repeat the study. One evaluation [5] was made by using Analytical Hierarchy Process (AHP) that we also use in our method. Therefore, probably in this case this study and our method are too similar to each other. There is a paper where the usability of just one system is analyzed [7]. On the other hand, our method is used in order to compare several systems with each other. Therefore, because of these reasons the evaluations that are described above are not suitable for our research. In the next section, we are going to describe the study [8] that we decided to repeat by using our method. We call this study as *control study*.

III. The Control Study

Martin at al. [8] conducted a research that compared usability of three e-learning platforms.

A. E-learning platforms

In that research "three of the most widely used open source Learning Management Systems" [8] were evaluated. Next, we will introduce them shortly.

Moodle is "a software package for producing internet-based courses and web sites" [18]. It is a software system that can be run on any computer supporting PHP and SQL database system. Tools of the system include assignment, chat, forums, glossary, quiz, calendar, blog, and course enrollment.

The development of the Sakai platform [19] began in 2004 as a joint project of Stanford, Michigan, Indiana, MIT, and Berkeley universities. It is a free Courseware Management System. The versions of software for Windows, Mac, and Linux systems are available. Tools include announcements, assignments, chat room, forums, messages, news, calendar, schedule, and wiki.

DotLRN [20] software is managed by the union of universities called dotLRN consortium. The system works on Linux operating systems. The main features of the system are assessment, e-mail, homework dropbox, file storage, forums, news, project manager, and survey.

B. The method of analysis

The evaluated platforms were tested with *default* layout and configuration. *Five* usability experts participated in the study. For each platform the experts carried out a series of tasks to test the main functions offered by the system. After completing the tasks each evaluator filled a data spreadsheet that was based on ten *Nielsen heuristics for heuristic evaluation* [9].

These heuristics (with short description) are as follows [10].

1. Visibility of system status – The system should always give users information about its current task in reasonable time and by using appropriate feedback.
2. Match between system and the real world – The system should speak the language of its users. It has to follow real-world conventions and present information in a natural and logical order.
3. User control and freedom – If users select some function of a system by mistake, then they need a clearly marked "emergency exit" to leave this state. Therefore, systems should support undo and redo operations.
4. Consistency and standards – The system should not force users to think whether different words and actions have the same meaning.
5. Error prevention – Eliminate error-prone conditions or ask confirmation from a user before a risky action.
6. Recognition rather than recall – Make objects, actions, and options visible. Instructions about how to use the system should be easily retrievable.
7. Flexibility and efficiency of use – Accelerators (the ways to make the work with the system faster) may speed up the interaction, so that the system can be efficiently used by experienced as well as inexperienced users.
8. Aesthetic and minimalist design – Dialogs should not contain irrelevant or rarely needed information.
9. Help users recognize, diagnose and recover from errors – Error messages should be expressed in plain language without using codes. They should define the problem precisely and suggest a solution.
10. Help and documentation – Information about the system should be easy to read and search.

More than 300 usability checkpoints have been created based on these heuristics [11]. Out of these checkpoints the ones that are irrelevant to e-learning platforms were dropped out. In case of each selected checkpoint c, the evaluators had to evaluate the systems in terms of c. They also had to evaluate the severity of the found problems.

In this paper, we are going to re-evaluate the systems that are mentioned in the control study by using our method. We also want to compare the results of the control study and our pattern-based evaluation. Therefore, in the next section we will shortly describe our method of analysis.

IV. Our Method of Analysis

The method that we are using in our analysis is based on using patterns as criteria for evaluation. For the analysis we use the Analytic Hierarchy Process [13] (AHP). According to this method we create a model with several levels. The highest level is the *goal* that we are going to achieve. At the second level are the *criteria* that we consider important in achieving our goal. At the third level of our model are *alternatives* – the systems that we are going to compare with each other. Our goal is to evaluate the usability of systems. We use patterns as criteria. A pattern is "a three-part rule, which expresses a relation between a certain context, a problem, and a solution" [12]. Our alternatives are the e-learning platforms, which we are going to evaluate. We compare elements of one level in respect to an element of higher level. Firstly, we compare all the patterns to see which of them are more important for usability. Next, we compare how well the selected patterns are implemented in the e-learning platforms. We calculate the relative importance of each pattern and the relative weight of each system in terms of each pattern. Finally, we combine the weights of criteria with the weights of systems and get the final results. The general method consists of 8 steps.

1. Selection of the systems that will be compared with each other and the patterns that will be the criteria for the analysis.
2. Pairwise comparisons of the selected criteria in order to determine their relative importance.
3. Writing and completing tests, evaluation of the results. The results are used as the basis of comparison of systems.
4. Pairwise comparison of the systems in terms of each criterion.
5. Calculation of the final weights for every system. This step allows us to determine the best system in terms of the selected criteria.
6. Identification and solving of the problems that were discovered during the testing (step 3).
7. Retesting of the system. We have to estimate how much the situation has changed after using the method.
8. Analysis of patterns. We have to evaluate the effectiveness of the patterns that were used in the analysis.

Although in this study and in [1] we used that method for evaluating usability, this method can be used to evaluate other aspects of software systems if other appropriate patterns are used as criteria. The next section will describe the organization of our study.

V. Organization of the Study

The main goal of our study is to get the results with our method and compare them with the results of the control study. Therefore, we do not have to complete all the steps of our method (see Section IV). More precisely, we do not have to complete the steps 6, 7 and 8. Next, we describe in more detail what we have done in our analysis (according to the steps of the method).

Selection of the Systems and the Patterns. The systems for the analysis are the same as in the control study (see Section III). Since the main criteria for evaluating the usability of systems were Nielsen's heuristics, the patterns that we use for our analysis are also related to these aspects. It was quite hard to find all the necessary patterns in one collection. Therefore, we used patterns from several sources. We have selected at most one pattern for each heuristic. In the following list the patterns are in the same order as the Nielsen's heuristics (see Section III) that they represent.

1. Sense of location (SL) [14] – Always provide visitors with information about their location at the site.

2. Natural metaphor (NM) [14] – Use metaphor that is natural to your target audience.

3. Go back to safe place (GBS) [14] – Provide a way to allow users to return if they get lost.

4. Follow standards (FS) [14] – Apply *de facto* and *de jure* standards.

5. Behavior constraint (BC) [17] – The system should prevent users from requesting hazardous actions. The system should anticipate these actions and allow only safe ones.

6. Recognition rather than recall (RRR) – For this heuristic we could not find a suitable pattern. We note that the study of existing heuristics may help to find candidate patterns.

7. Shortcut box (SB) [15] – Let users select important locations and functions from a combo box.

8. Keep it simple (KISS) [14] – Remove from the design things that are not central to the message you wish to express.

9. Input error message (IEM) [15] – If users have entered input that has invalid format, then tell them that there is a problem. In addition, explain them how to solve this problem and where did this problem arise.

10. Inline help box. (IHB) [16] – If users are about to perform an interaction, then they need a helpful introduction.

Comparison of the selected patterns. While comparing the selected patterns with each other, we tried to evaluate how important each pattern could be for the students who use these systems (see Section VI A).

Creating and using the tests. We performed the same tasks that were mentioned in the control study. These tasks were the following.

1. Register in the platform.
2. Enroll to a course.
3. Download a document.
4. Post a message to a forum.
5. Add a personal event to the calendar.

Calculation of overall results. After completing the tests, we compared the use of each pattern in each system. The evaluation of systems was made by *one* usability expert. We calculated the final result by combining weights that characterize the relative importance of patterns and weights that characterize the use of patterns in the systems.

Next, we are going to present the result of our study.

VI. The Results

In this section, we will firstly provide the results of criteria comparison. Then, we will describe the problems that were discovered while completing the test tasks. After that, we will shortly discuss how well each pattern is implemented in the different systems. And finally the overall results will be described. Numerical results of our study are shown in Table I.

A. Comparison of Criteria

Here we describe the results of comparison of patterns. The weight of each pattern can be seen in the row "Pattern Weight" (see Table I).

In our comparison the highest weight (0,193) got "Follow Standards" (FS) pattern. We think that if a user can accomplish some tasks using his/her previous experience with other e-learning systems or other software systems in general, then it makes learning a new system much easier.

TABLE I
WEIGHTS OF PATTERNS AND E-LEARNING SYSTEMS

	SL	NM	GBS	FS	BC	RRR	SB	KISS	IEM	IHB	Sum
Pattern Weight	*0,072*	*0,031*	*0,096*	*0,193*	*0,157*	*0,103*	*0,037*	*0,055*	*0,132*	*0,124*	
Moodle	0,4	0,540	0,2	0,2	0,4	0,429	0,333	0,429	0,143	0,429	0,318
Sakai	0,2	0,297	0,4	0,4	0,4	0,143	0,333	0,143	0,429	0,429	**0,347**
dotLRN	0,4	0,163	0,4	0,4	0,2	0,429	0,333	0,429	0,429	0,143	0,335

The second highest weight (0,157) got the pattern "Behavior Constraint" (BC) because it is quite important to assure that a user will not do something that would lead to unwanted consequences.

Next two patterns that got quite even result are "Input error message" (IEM) (0,132) and "Inline help box" (IHB) (0,124). IEM is important because each system should always give a feedback to a user about what is going on, especially in case of an error.

IHB is somewhat similar because if a user has some problem with a system (for example if he/she does not know how to use some function), then it should be possible to find help material quickly and easily.

"Recognition rather than recall" (RRR) is quite important heuristic. It ensures that the feedback that a system gives to a user is complete for accomplishing his/her task.

B. Testing of the Systems.

The control study states that in the testing of each system "current stable version with default configuration" was used. The tests took place in January 2008.

In our study, we have used the following versions of the e-learning platforms:

Moodle 1.9.2 – latest official release (11th July 2008);

Sakai 2.5.2 – demo version for Linux;

Virtual machine with dotLRN (version 1.3.1).

While accomplishing the tasks mentioned in section V, we also tried to do something wrong in order to see how the systems respond to these actions. Additionally we tried to delete some data like forum posts or calendar events to see if the system warns us about possibly harmful action.

Register to the platform – Moodle in its registration form has only one field for password input (but it has an option to make the password visible). On the other hand, it has two fields for e-mail address. When these e-mails do not match, the system says "Wrong e-mail address". This message does not describe exactly how to solve the problem. It should say that e-mail addresses do not match. Sakai has no problem with registration process. But there is a field in the registration form named "type" that a user cannot edit. Therefore it probably should be hidden. DotLRN does not have any problems in this area.

Register to the course – No serious problems were discovered during the testing. The only shortcomings were imprecise terms for enrolling a course.

In Sakai this function is called "Membership" and in dotLRN a user can join "classes" and "communities". It is not quite clear how these two functions are related to each other. It seems that if a user joins "class", then he/she automatically joins "community" as well.

Post a message to a forum – We discovered some problems in the Moodle system while posting a forum message with an attachment. If a user tries to send a forum post with a too big attachment file, then the system error is not firstly visible. The reason for this is that the error message is located too low and the user needs to scroll down the page in order to see it. In addition, if a user decides for some reason to cancel the posting, then there is no explicit way to do that. He/she has to select another function in the system.

In the Sakai system the maximum allowed size of an attachment to a forum post is not clearly stated. In addition, if a user adds an attachment, then a new form opens up and it does not say that this is for adding an attachment to a forum post.

Download document – did not have any problems.

Add a personal event to a calendar – Sakai has two functions "Calendar" and "Schedule", so it was a bit confusing. We found out that calendar shows events that are connected to a course. A user can add his/her personal events to the schedule. By default the system offers "--" hours and "00" minutes as the starting time for an event. It looks a bit strange and unclear. In addition, it may be a bit inconvenient that a user cannot add an event, which has no special time for it.

C. Evaluating the Use of Patterns in the Systems.

In this section, we introduce the results of pairwise comparison of systems in terms of different patterns. Rows in Table I correspond to systems that are tested and columns correspond to patterns. Cells contain the results of pairwise comparisons. Table II and III show the results of the pairwise comparisons of systems in respect to the patterns "Input error message" and "Inline help box", respectively. Next, we will describe how well the systems conform to each pattern. We will shortly mention the problems that were already described in the previous subsection in order to improve readability.

Sense of location – We have found only one problem connected to this pattern. If a user of the Sakai system adds an attachment, then the system does not give enough information on what is going on.

Natural Metaphors – There were some unclear words that were used for the enrollment function in the Sakai and dotLRN systems.

TABLE II
COMPARISONS AND WEIGHTS OF SYSTEMS IN THE CONTEXT OF "INPUT ERROR MESSAGES" PATTERN

	Moodle	Sakai	.LRN	Weight
Moodle	1	1/3	1/3	0,143
Sakai	3	1	1	0,429
dotLRN	3	1	1	0,429

TABLE III
COMPARISONS AND WEIGHTS OF SYSTEMS IN THE CONTEXT OF "INLINE HELP BOX" PATTERN

	Moodle	Sakai	.LRN	Weight
Moodle	1	1	3	0,429
Sakai	1	1	3	0,429
dotLRN	1/3	1/3	1	0,143

Go Back To Safe Place – There is no explicit way to cancel forum posting in the Moodle system.

Follow Standards – Nothing unusual except the fact that the Moodle system uses one field for password and two fields for e-mail address in the registration form.

Behavior Constraint – If a user drops himself/herself from a community or a class in dotLRN, then he/she does not get warnings and the system does not ask confirmation.

Recognition rather than recall – In the Sakai system there is no information on the maximum size of an attachment that can be added to a forum post. In addition, the use of schedule is a bit confusing.

Shortcut box – All the systems seem to have enough means to enable experienced user to access various functions rather quickly.

Keep it simple – Sakai system has a visible field in the registration form that a user cannot edit.

Input error message – The Moodle system has two faults in case of this pattern. The main problem is that one of its error messages is not clearly visible. In addition, one error message defines the problem imprecisely.

Inline help box – User interfaces of the systems Moodle and Sakai provide links to help material near to almost each form element. On the other hand, in the dotLRN system it was somewhat hard for us to find help documentation.

D. The Final Results

As we can see from the last column of Table I (Sum), the best result got the Sakai system. On the second place we find dotLRN. Moodle got the third place. The differences between the weights of these three systems are quite small. It shows that these three systems are quite even in terms of usability.

Next, we will shortly describe the results of the control study.

VII. RESULTS OF THE CONTROL STUDY

Overall results. According to the results of the control study the best system is dotLRN that is compliant with 78% percent of checkpoints. A bit behind is Sakai with 77%. Moodle is at the last place with 68% of compliance.

Results for heuristics. Moodle achieved the best results in case of the heuristic "Help users recognize, diagnose and recover from errors". The worst results for this system were in case of the heuristic "Flexibility and efficiency of use". The best results for Sakai and dotLRN was in case of the heuristic "Aesthetic and minimalist design". The worst heuristic for these two system was also "Flexibility and efficiency of use".

In case of heuristics "help users recognize, diagnose and recover from errors" and "aesthetic and minimalist design" all the systems got around 80% of checkpoints. The systems Sakai and dotLRN scored around 80% also in case of heuristics number 1, 4, and 6.

Severity of problems. The control study also paid attention to the severity of the found problems (see Section III). 90% of the found problems had either low or medium severity.

Next, we will compare and discuss the results of our study with the results of the control study.

VIII. COMPARISON AND DISCUSSION

Methods. We tried to make our study as similar as possible to the control study, but still there were differences. The most important of them in our opinion are the following.

- The report of the control study does not specify exact versions of the systems.
- Experts in the control study used a questionnaire and the systems were evaluated "independently". Our method uses AHP and experts should compare systems pairwise.
- In case of the control study, experts used about 20 checkpoints for each heuristic [8]. On the other hand, we used in our study the descriptions of *nine* patterns and *one* heuristic. Each pattern corresponded to exactly one heuristic.
- "Heuristic" and "pattern" are different concepts. In our case heuristic is a general usability principle and pattern is a more thorough and structured description of a problem and its solution. Therefore, the results of comparisons of systems for specific heuristic and pattern may differ.
- The evaluations in the control study and our study were made by *five* and *one* experts, respectively.

Overall results. While the results of our study are not quite the same as the results of the control study, they are still quite similar. The DotLRN system had a very little advantage over the Sakai system in the control study. In our study Sakai has a little advantage over dotLRN. The reason for that could be that we used the version of the Sakai system was somewhat newer than the version of the dotLRN system. In both studies the Moodle system got the third place. However, we cannot directly compare the numerical results of our study with the results of the control study (see Section VII). The reason is that in the control study the systems were tested against checkpoints separately. While in our study we evaluated how much one system is better in comparison to other systems.

Results for particular heuristics. If we compare the results of the e-learning systems in the context of specific heuristics, then we also will notice difference in the results. In the control study Moodle had its best results in case of the heuristics “help users recognize, diagnose and recover from errors”. On the other hand, in our study Moodle got its worst score in case of the pattern IEM that should correspond to “help users...” heuristic. The reason is that in our testing we took into account not only what error messages say but also how they appear. One of the error messages was not very visible (see Section VI).

These two examples show that the results of the two studies are difficult to compare directly. The reasons are the differences in evaluation criteria and in the comparison method of systems.

Severity of the found problems. We have not explicitly evaluated the severity of the found problems. However, we can determine the severity of problems based on the relative importance of the patterns (see row "pattern weight" in Table I). The higher is the weight of a pattern, the more severe are the problems that were detected in connection with this pattern. It means that these problems must be fixed first. In our case, the most severe problems were detected in connection with pattern "Follow standards" in the Moodle system.

The method that was used in the control study is good for giving approximate level of usability of systems. The advantage of our method of analysis is that patterns describe the possible solutions to the problems, context, forces and the results of using the solution. Heuristic says just what should be done (See for example heuristic number 7 in the section III). Therefore, the use of patterns simplifies the improvement of systems (see step 6 of our proposed method).

Another advantage of our proposed method is that it is easier to compare objects (systems, patterns etc.) pairwise than to consider all them together.

We conclude that our proposed method can be used in order to compare different systems. The method gives similar results as one existing method but at the same time simplifies the evaluation and provides guidelines for improving systems.

The results of our study could also be interesting to educational institutions, which have to select their e-learning platform. In addition, this kind of studies might help to identify new candidate patterns and also to evaluate how widespread is the use of different patterns. If there is more than one evaluation that uses the same set of patterns as criteria in the same context (for instance e-learning), then it would help us to determine the relative importance of different patterns based on opinions of different experts.

IX. FUTURE WORK

We have the following possible directions for development.

- Generalize our proposed method and find criteria for selecting systems and patterns.
- Research the possibility of using group judgments.
- Create a software system for using the generalized method with different types of patterns. For calculating the AHP weights, we used *Open Office Calc*.

X. CONCLUSIONS

We found a study of usability of e-learning platforms. In that study three systems were analysed using ten heuristics. We re-evaluated the usability of these three systems by using AHP method. We used nine usability patterns and one heuristic as the criteria. We provided the numerical results of comparative analysis of usability of these three systems. In addition, we compared our results with the results of the earlier study and discussed the differences.

REFERENCES

[1] D. Zub and E. Eessaar, “Pattern-based usability evaluation of e-learning systems,” in: Innovative Techniques in Instruction Technology E-learning, E-assessment and Education, M. Iskander, Ed. Springer 2008 pp. 117-122.

[2] G. Bafoutsou and G. Mentzas, “A comparative analysis of web-based collaborative systems,” In: Proceedings of 12th international Workshop on Database and Expert Systems Applications, A. M. Tjoa, R. R. Wagner, Eds. IEEE Computer Society Press 2001, pp.496-500.

[3] A. Karoulis, “Usability Evaluation of the LEGO interface,” In: Proceedings of the Sixth IEEE International Conference on Advanced Learning Technologies, IEEE Computer Society Press 2006, pp. 1052-1054.

[4] F. Colace, M. De Santo, and M. Vento, “Evaluating On-line Learning Platforms: a Case-Study,” In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences, IEEE 2003, pp. 154-162

[5] E.W.T. Ngai and E.W.C. Chan, “Evaluation of knowledge management tools using AHP,” *Expert Systems with Applications*, vol. 29, issue 4, pp. 889-899, 2005.

[6] M. Kljun, J. Vicic, B. Kavsek, and A. Kavcic, “Evaluating Comparisons and Evaluations of Learning Management Systems,” In: 29th International Conference on Information Technology Interfaces 2007, IEEE 2007, pp. 363-368.

[7] J. Komarkova, M. Novak, R. Bilkova, O. Visek, and Z. Valenta, “Usability of GeoWeb Sites: Case Study of Czech Regional Authorities Web Sites,” LNCS vol. 4439, pp 411-423, 2007.

[8] L. Martin, D. Roldán Martinez, O. Revilla, M.J. Aguilar., O.C. Santos, and J. Boticario. “Usability in e-Learning Platforms: heuristics comparison between Moodle, Sakai and dotLRN,” In: Sixth International Conference on Community based environments. Guatemala, 12–16, February 2008.

[9] J. Nielsen, “Heuristic evaluation,” Retrieved September 30, 2008 from http://www.useit.com/papers/heuristic/

[10] J. Nielsen “Ten Usability Heuristics,” Retrieved September 30, 2008 from http://www.useit.com/papers/heuristic/heuristic_list.html

[11] D. Pierotti “Heuristic Evaluation – A System Checklist,” Retrieved September 30, 2008 from http://www.stcsig.org/usability/topics/articles/he-checklist.html

[12] C. Alexander, M. Silverstein, S. Angel, S. Ishikawa, and D. Abrams, *The Oregon Experiment*, Oxford University Press, New York, 1975.

[13] T. L. Saaty, “How to Make a Decision: The Analytic Hierarchy Process,” *European Journal of Operational Research*, vol. 48, no. 1, pp. 9-26, 1990.

[14] Web Usability Patterns. Retrieved September 30, 2008, from http://www.trireme.com/WU

[15] M. Welie, “Pattern Library,” Retrieved September 30, 2008 from http://www.welie.com/patterns/index.php

[16] UI Patterns User Interface. User Interface Design Pattern Library. Retrieved September 30, 2008 from http://ui-patterns.com/

[17] M. Mahemoff, A. Hussey, and L. Johnston “Pattern-based Reuse of Successful Designs: Usability of Safety Critical Systems,” In 2001 Australian Software Engineering Conference: Proceedings, 2001. pp. 31-39

[18] Moodle. Retrieved September 30, 2008 from http://www.moodle.org

[19] Sakai. Retrieved September 30, 2008 from http://www.sakaiproject.org/portal

[20] dotLRN. Retrieved September 30, 2008 from http://www.dotlrn.org

The Effect of Pressure in Two-Dimensional Luikov Drying Problem

João Batista Furlan Duarte
Universidade de Fortaleza - UNIFOR, CCT-MIA/NTC, Fortaleza - CE furlan@unifor.br
Júlio Wilson Ribeiro
Universidade Federal do Ceará - UFC, Departamento de Computação, jwilson@lia.ufc.br
Plácido Rogério Pinheiro
Universidade de Fortaleza - UNIFOR, CCT-MIA, Fortaleza – CE, placido@unifor.br
Carlos Almir Monteiro de Holanda
Universidade Federal do Ceará - UFC, Departamento de Engenharia Metalúrgica e de Materiais, almir@ufc.br

Abstract
Multidimensional drying in capillary porous media is analytically solved for the associated temperature, moisture and pressure content distributions. Luikov's model with linear transport coefficients and two-dimensional plate geometry is adopted for description of the simultaneous heat, mass and pressure transfer phenomena. The generalized integral transform technique (G.I.T.T.) is applied to the problem and the automatically global error-controlled solution of the coupled partial differential equations is used to achieve the solutions. The convergence behavior of the proposed eigenfunction expansions is illustrated.

1. Introduction

The system of equations proposed by Luikov[1] is by far the most frequently adopted in the study of drying phenomena in capillary porous media with various applications in the engineering and applied sciences. The integral transform method[2,3] has been successfully utilized in the hybrid numerical-analytical solution of such problems, for both the linear[4,5,6] and non-linear versions[7,5], offering the attractive feature of automatic global error control in the final results. Both applications previously considered[4,5,6,7] where the interest in studying multidimensional situations are ever increasing, as demonstrated by the finite element method numerical solution in Ferguson and Lewis[8]. Therefore, the present contribution advances the integral transform methodology to be applicable in multidimensional drying problems, such as the one formulated in Lewis et al.[9] and Thomas et al.[10], and demonstrates another attractive feature of this class of hybrid method, i.e., the just very mild increase in computational effort for increased number of dimensions in the problem (independent variables). Essentially, it is reconfirmed that the overall computational cost in implementing the one-dimensional simulation is exactly comparable to that of solving the two-dimensional problem here proposed.

2. General Solution Using the Integral Transform Technique

The simultaneous heat, mass and pressure transfer in porous media can be expressed by a set of three parabolic equations, defined in the finite region $\boldsymbol{V}$ with boundary surface $\boldsymbol{S}$, coupled through both diffusion equations and boundary conditions[11,6]:

$$w_k(\mathrm{x})\frac{\partial\theta_k(\mathrm{x},t)}{\partial t}+L_k\theta_k = P_k(\mathrm{x},t,\theta_1,\theta_2,\theta_3), \quad \mathrm{x}\in \mathrm{V},\ t>0,\ k=1,2,3 \tag{1,2,3}$$

and initial and boundary conditions given respectively by:

$$\theta_k(\mathrm{x},0)=f_k(\mathrm{x}), \mathrm{x}\in \mathrm{V},\ k=1,2,3 \tag{3,4,5}$$

$$B_k\theta_k(\mathrm{x},t)=\phi_k(\mathrm{x},t,\theta_1,\theta_2), \mathrm{x}\in \mathrm{S}, k=1,2,3 \tag{6,7,8}$$

where the equation and boundary operators are written as:

$$L_k \equiv -\nabla.K_k(\mathrm{x})\nabla + d_k(\mathrm{x}),\ k=1,2,3 \tag{9,10,11}$$

$$B_k \equiv [\gamma_k(\mathrm{x})+\delta_k K_k(\mathrm{x})\frac{\partial}{\partial\eta}],\ k=1,2,3 \tag{12,13,14}$$

$\gamma_k(\boldsymbol{x})$ and $\delta_k(\boldsymbol{x})$ are prescribed boundary condition coefficients. η is the outward drawn normal to surface S. The non-linear source terms, P_k and ϕ_k, may incorporate any coupling between the two potentials.

After, the approach deals with the selection of filtering solutions to minimize the non-homogeneities effects:

$$\theta_k(\mathrm{x},t)=\theta_{ks}(\mathrm{x},t)+\theta_{kh}(\mathrm{x},t) \tag{15}$$

where the quasi-steady solutions θ_{ks}'s, include the consideration of characteristic representations of the equation and boundary source terms.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_43,

The resulting formulation for the filtered potential $\theta_{kh}(\mathrm{x},t)$, is obtained from the solution of the following easier problems:

$$L_k\theta_{ks} = P_{ks}(\mathrm{x},t), \mathrm{x} \in \mathrm{V}, k = 1,2,3 \quad (16,17,18)$$

$$B_k\theta_{ks} = \phi_{ks}(\mathrm{x},t), \mathrm{x} \in \mathrm{S}, k = 1,2,3 \quad (19,20,21)$$

Following the formalisms, the homogeneous system[6,2,3,12] could be solved exactly through the integral transform technique, since the solution of the following independent auxiliary decoupled eigenvalue problems (*k=1,2*) of Sturm-Liouville type is available,

$$L_1\Psi_i(\mathrm{x}) = \mu_i^{\,2} w_1(\mathrm{x})\Psi_i(\mathrm{x}), \mathrm{x} \in \mathrm{V}, i = 1,2... \quad (22)$$

$$L_2\Gamma_i(\mathrm{x}) = \lambda_i^{\,2} w_2(\mathrm{x})\Gamma_i(\mathrm{x}), \mathrm{x} \in \mathrm{V}, i = 1,2... \quad (23)$$

with boundary conditions:

$$B_1\Psi_i(\mathrm{x}) = 0, \mathrm{x} \in \mathrm{S}, i = 1,2... \quad (24)$$

$$B_2\Gamma_i(\mathrm{x}) = 0, \mathrm{x} \in \mathrm{S}, i = 1,2... \quad (25)$$

Through the associated orthogonality property of the eigenfunctions, we define the integral transform pair as follows:

Transform:

$$\bar{\theta}_{1,i}(t) = \int_V w_1(\mathrm{x})\frac{\Psi_i(\mathrm{x})}{N_i^{1/2}}\theta_{1h}(\mathrm{x},t)dv \quad (26)$$

Inverse:

$$\theta_{1h}(\mathrm{x},t) = \sum_{i=1}^{\infty}\frac{\Psi_i(\mathrm{x})}{N_i^{1/2}}\bar{\theta}_{1i}(t) \quad (27)$$

and,

Transform:

$$\bar{\theta}_{2,i}(t) = \int_V w_2(x)\frac{\Gamma_i(x)}{M_i^{1/2}}\theta_{2h}(x,t)dv \quad (28)$$

Inverse:

$$\theta_{2h}(\mathrm{x},t) = \sum_{i=1}^{\infty}\frac{\Gamma_i(\mathrm{x})}{M_i^{1/2}}\bar{\theta}_{2i}(t) \quad (29)$$

The normalizations integrals are:

$$N_i = \int_V w_1(\mathrm{x})\Psi_i^2(\mathrm{x})dv \quad (30)$$

$$M_i = \int_V w_2(\mathrm{x})\Gamma_i^2(\mathrm{x})dv \quad (31)$$

Using the integral transform methodology for the homogeneous system, after truncation to a sufficient order *N*, for the desired convergence, we obtain a transform constant coefficient ordinary differential equations system:

$$\frac{d\mathrm{Y}(t)}{dt} + \mathrm{A}_{2N,2N}\,\mathrm{Y}(t) = 0_{2N,1} \quad (32)$$

where

$$\mathrm{Y}(t) = \{\bar{\theta}_{11}(\tau), \bar{\theta}_{12}(\tau)...\bar{\theta}_{21}(\tau), \bar{\theta}_{22}(\tau)....\bar{\theta}_{2N}(\tau)\}^T \quad (33)$$

The initial transform conditions are obtained from the filtered system. Equation (32) can be readily solved through matrix eigenvalue analysis or through well-established algorithms. Temperature and moisture potentials are computed from the explicit analytic inverse formulae on (27 and 29).

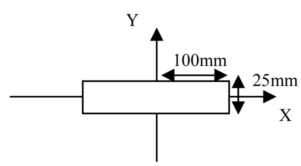


Figure 1 . Evolution of temperature profiles during the process.

3. Application

We consider the heat, mass and pressure balance equations written in dimensionless form, for a symmetric plate geometry as depicted in Ferguson and Lewis[8], Fig. (1) subjected to uniform prescribed boundary temperatures, moisture and pressure contents, and evaluated from uniform initial distributions[4,5]. The transport coefficients are assumed constant and the problem formulation according to Luikov's theory[1] is given by Ferguson and Lewis[8]:

$$\frac{\partial\Theta_1(X,Y,\tau)}{\partial\tau} = K_{11}\nabla^2\Theta_1(X,Y,\tau) + K_{12};$$
$$\nabla^2\Theta_2(X,Y,\tau)K_{12} + K_{13}\nabla^2\Theta_3(X,Y,\tau)$$
$$0<X<1,\ 0<Y<1;\ \tau>0 \quad (34)$$

$$\frac{\partial\Theta_2(X,Y,\tau)}{\partial\tau} = K_{21}\nabla^2\Theta_1(X,Y,\tau) + K_{22}\nabla^2\Theta_2(X,Y,\tau) + K_{23}\nabla^2\Theta_3(X,Y,\tau);$$

$0<X<1, 0<Y<1;\ \tau>0$ (35)

$$\frac{\partial\Theta_3(X,Y,\tau)}{\partial\tau} = K_{31}\nabla^2\Theta_1(X,Y,\tau)+K_{32}\nabla^2\Theta_2(X,Y,\tau)+ K_{33}\nabla^2\Theta_3(X,Y,\tau);\quad 0<X<1;\quad 0<Y<1;\tau>0 \quad (36)$$

with initial conditions

$$\Theta_1(X,Y,0)=\Theta_2(X,Y,0)=\Theta_3(X,Y,0)=0;\quad 0<X<1,\ 0<Y<1 \quad (37,38,39)$$

and boundary conditions

$$\frac{\partial\Theta_1(0,Y,\tau)}{\partial X}=0;\quad \frac{\partial\Theta_1(X,0,\tau)}{\partial Y}=0;\quad \tau>0 \quad (40,41)$$

$$\frac{\partial\Theta_2(0,Y,\tau)}{\partial X}=0;\quad \frac{\partial\Theta_2(X,0,\tau)}{\partial Y}=0;\quad \tau>0 \quad (42)$$

$$\frac{\partial\Theta_3(0,Y,\tau)}{\partial X}=0;\quad \frac{\partial\Theta_3(X,0,\tau)}{\partial Y};\quad \tau>0 \quad (43)$$

$$\Theta_1(1,Y,\tau)=\Theta_2(1,Y,\tau)=\Theta_3(1,Y,\tau)=1;\ \tau>0 \quad (44,45,46)$$

$$\Theta_1(X,1,\tau)=\Theta_2(X,1,\tau)=\Theta_3(X,1,\tau)=1;\ \tau>0 \quad (47,48,49)$$

where the K´s represent:

$$K_{11}=k_q+\varepsilon\lambda\delta';\quad K_{12}=\varepsilon\lambda k_m;\quad K_{13}=\varepsilon\lambda k_p \quad 50,51,52)$$

$$K_{21}=\delta k_m;\quad K_{22}=k_m;\quad K_{23}=k_p \quad (53,54,55)$$

$$K_{31}=-\varepsilon\lambda k_m;\ K_{32}=-\varepsilon k_m;\ K_{33}=k_p(1-\varepsilon) \quad (56,57,58)$$

and θ_1 is the dimensionless temperature distribution, θ_2 is the dimensionless moisture content distribution, θ_3 is the dimensionless pressure distribution.

Without loss of generality, using the formalisms of the *integral transform*[6] method the solution for the system of . (34,35,36) is now proposed in terms of auxiliary problems, expressed by three pairs of easily available decoupled eigenfunction expansions of Sturm-Liouville problems, for the temperature, moisture and pressure potentials (k = 1,2,3):

$$\frac{d^2\Psi_{ki}(X)}{dX^2}+\mu_{ki}^2\Psi_{ki}(X)=0,\qquad X\in\mathbf{V} \quad (59)$$

$$\frac{d\Psi_{ki}(0)}{dX}=0,\quad \Psi_{ki}(1)=0 \qquad X\in\mathbf{S} \quad (60,61)$$

$$\frac{d^2\Gamma_{kj}(Y)}{dY^2}+\lambda_{kj}^2\Gamma_{kj}(Y)=0,\qquad Y\in\mathbf{V} \quad (62)$$

$$\frac{d\Gamma_{kj}(0)}{dY}=0,\quad \Gamma_{kj}(1)=0 \qquad Y\in\mathbf{S} \quad (63,64)$$

These auxiliary problems permit the definition of the integral transform pairs that are necessary for the solution of the homogeneous problem:

Inverse,

$$\Theta_{kh}(X,Y,\tau)=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{N_{ki}^{1/2}M_{kj}^{1/2}}\Psi_{ki}(X)\Gamma_{kj}(Y)\overline{\Theta}_{kij}(\tau) \quad (65)$$

Transform,

$$\overline{\Theta}_{kij}(\tau)=\int_0^1\int_0^1\frac{\Psi_{ki}(X)\Gamma_{kj}(Y)}{N_{ki}^{1/2}M_{kj}^{1/2}}\Theta_{kh}(X,Y,\tau)dXdY \quad (66)$$

The normalizations integrals are,

$$N_{ki}=\int_0^1\Psi_{ki}^2(X)dX \quad (67)$$

$$M_{kj}=\int_0^1\Gamma_{kj}^2(Y)dY \quad (68)$$

The problem now is to find numerically the eigenvalues (μ_{ki}, and λ_{kj}), eigenfunctions (Ψ_{ki}, and Γ_{kj}) and norms (N_{ki}, and M_{kj}).

The next step is to find the ordinary differential equation transform system. Using the *transform concept* in (34-58) and the auxiliary problems (59-64,67,68), after truncation to a sufficient order (i =1 **...** I, and j = 1 **...** J) for the desired convergence, we obtain,

$$\frac{dY(\tau)}{d\tau}+A_{2N,2N}Y(\tau)=0_{2N,1} \quad (43)$$

where,

$$Y(\tau)=\{\overline{\Theta}_{11}(\tau)...\overline{\Theta}_{1N}(\tau)\quad \overline{\Theta}_{21}(\tau)...\overline{\Theta}_{2N}(\tau)\}^T \quad (44)$$

The initial transform conditions are similarly obtained applying the integral transform concept to the initial conditions on the homogeneous problem, resulting:

$$Y(0)=\overline{F}(\tau) \quad (45)$$

Now, this initial value problem can be solved through matrix eigenvalue analysis or scientific libraries. Initial value problem solvers with local error control schemes are employed for solving the truncated version of the transformed initial value problem. An adaptive procedure is utilized to automatically reduce, along the integration path, the truncation orders required for a certain user-prescribed accuracy yielding, as a by-product, a global error estimator.

At this point, it is possible written the complete solution to the original problem. Using inversion formulae, temperature and moisture potentials can now be numerically obtained as:

$$\Theta_k(X,Y,\tau) = \Theta_{ks}(X,Y) + \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{\Psi_{ki}(X)\Gamma_{kj}(Y)}{M_{ki}^{1/2} N_{kj}^{1/2}} \overline{\Theta}_{kij}(\tau)\ ;$$

$\tau > 0$ (46)

where $\boldsymbol{\Theta_{ks}}$ are the steady-state solutions, $\boldsymbol{\Psi_{ki}}$ and $\boldsymbol{\Gamma_{kj}}$ are the normalized eigenfunctions, and $\overline{\Theta}_{kij}$ represent the transformed potentials, obtained from numerical solution of the resulting ordinary differential system, after the completion of the integral transformation process.

4. Results And Discussion

The Luikov problem as proposed above is now solved using the integral transform technique. The numerical results make it possible an inspection of the overall convergence behavior for the proposed eigenfunction expansions. The governing parameters, according to the data in Lewis[9], and Cunha et al.[13], assume the following values: $\rho_0 = 1170{,}0$ $Kg.m^{-3}$, $c_q = 1.400{,}0$ $J.Kg^{-1}.^oK^{-1}$, $c_m = 0{,}03$ $Kg.Kg^{-1}.^oM^{-1}$, $c_p = 0{,}05$ $Kg.Kg^{-1}.Pa$, $\varepsilon = 0{,}3$, $\lambda = 2{,}3.10^6$ $J.Kg^{-1}$, $\Delta = 0{,}67$ $^oM.^oK^{-1}$, $k_q = 576{,}0$ $J.h^{-1}.m^{-1}.^oK^{-1}$, $k_m = 3{,}0.10^{-6}$ $Kg.h^{-1}.m^{-1}.^oM^{-1}$, $K_p = 1{,}5.10^{-6}$ $Kg.h^{-1}.m^{-1}.Pa^{-1}$, and the truncation orders, *N*, were taken less or equal to 9, for temperature, moisture and pressure. The computer program was implemented on *Mathemtica®* software[14], on a Pentium 700 *MHz* microcomputer with 256 Mb of memory RAM, and a typical run took less than *5* minutes of CPU time.

Tables (1,2) below illustrates the convergence behavior of the two expansions (different *N*'s) for temperature ($\boldsymbol{\Theta_1}$), moisture ($\boldsymbol{\Theta_2}$) and pressure potential ($\boldsymbol{\Theta_3}$), obtained at the plate centerlines ($Y = 0.5$) and different *X* positions. Since the heat, mass and pressure transfer processes have, for this problem, markedly different time constants, the values of dimensionless time considered in each case, are different. The convergence characteristics are, in both potentials, quite evident, with full convergence to four digits to moisture and pressure distribution and three digits to temperature distribution being achieved at *N* as low as *12*. Such results open up broad perspectives for extension of this approach into even more involved coupled parabolic problems.

Figures (2-5) show drying process, and the temperature, moisture and pressure distributions are obtained with the converged values. As expected, the epoxy have a low thermal inertia spending about *0,1* dimensionless time to achieve the thermal equilibrium in the most deep layer, and are needed more than *1500* dimensionless times steps to the porous media to meet the moisture equilibrium. This characterize a very right mass inertia. The same can be said for the pressure, which has a very right inertia achieving the equilibrium state before the moisture content, with *1400* dimensionless time steps. The different time equilibrium make the difference in O.D.E convergence number, for moisture, because the negative pressure work over the material in the same time, carriage to deep inside a considerable quantity of humidity mass, allied to thermo-gradient effect. Such process create some difficulty to the numerical convergence, and of course, to the real drying process, as it take place after *300* dimensionless times steps, as can be seen in Fig (4). The drying process can be observed, when the pressure potential reached a half value of the prescribed boundary and the temperature is established over the material ($\tau > 300$).

Table 1 . Convergence behavior of temperature, Θ_1 moisture, Θ_2 and pressure, Θ_3 expansions.

$\Theta_1(X,0.\tilde{0},0.01)$				
X/N	3	6	9	12
0.0	0.1494	0.1790	0.1790	0.1785
0.2	0.1931	0.1781	0.1790	0.1785
0,4	0.1990	0.1783	0.1789	0.1787
0,6	0.1417	0.1845	0.1841	0.1841
0,8	0.3574	0.3215	0.3218	0.3218
1,0	1.0000	1.0000	1.0000	1.0000
	$\Theta_2(X,0.0,150)$			
X/N	3	6	9	12
0.0	3.3525	3.2007	3.2030	3.2030
0.2	3.1346	3.2072	3.2057	3.2057
0,4	3.1465	3.2445	3.2440	3.2440
0,6	3.6697	3.4551	3.4575	3.4576
0,8	3.2215	3.3953	3.3928	3.3928
1,0	1.0000	1.0000	1.0000	1.0000
	$\Theta_3(X,0.0,250)$			
X/N	3	6	9	12
0.0	0.3131	0.3220	0.3220	0.3220
0.2	0.3264	0.3216	0.3216	0.3216
0,4	0.3256	0.3213	0.3213	0.3213
0,6	0.3375	0.3480	0.3480	0.3480
0,8	0.5437	0.5349	0.5349	0.5349
1,0	1.0000	1.0000	1.0000	1.0000

Table 2 . Convergence behavior of temperature, Θ_1 moisture, Θ_2 and pressure, Θ_3 expansions

$\Theta_1(X,0.0,0.025)$				
X/N	3	6	9	12
0.0	0.5622	0.5636	0.5643	0.5638
0.2	0.5661	0.5642	0.5645	0.5642
0,4	0.5695	0.5680	0.5685	0.5684
0,6	0.5996	0.6012	0.6016	0.6016
0,8	0.7384	0.7359	0.7353	0.7353
1,0	1.0000	1.0000	1.0000	1.0000

$\Theta_2(X,0.0,600)$				
X/N	3	6	9	12
0.0	2.1274	2.1270	2.1270	2.1270
0.2	2.1193	2.1195	2.1195	2.1196
0,4	2.0718	2.0720	2.0720	2.0720
0,6	1.9079	1.9075	1.9075	1.9075
0,8	1.5412	1.5419	1.5419	1.5419
1,0	1.0000	1.0000	1.0000	1.0000
$\Theta_3(X,0.0,400)$				
X/N	3	6	9	12
0.0	0.5697	0.5709	0.5709	0.5709
0.2	0.5720	0.5713	0.5713	0.5713
0,4	0.5787	0.5782	0.5782	0.5782
0,6	0.6198	0.6211	0.6211	0.6211
0,8	0.7590	0.7579	0.7579	0.7579
1,0	1.0000	1.0000	1.0000	1.0000

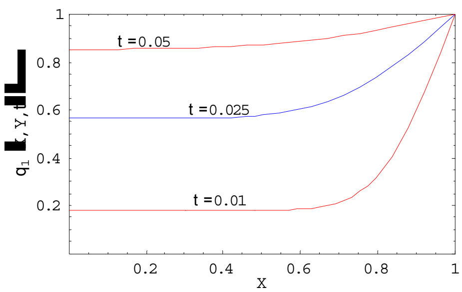


Figure 2. Evolution of temperature profiles during the process.

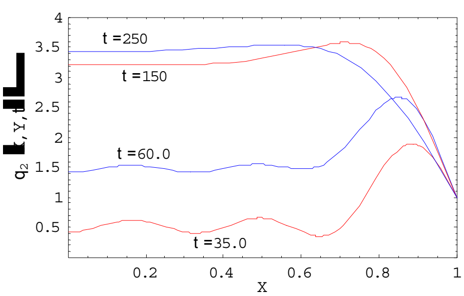


Figure 3. Evolution of moisture profiles during the process

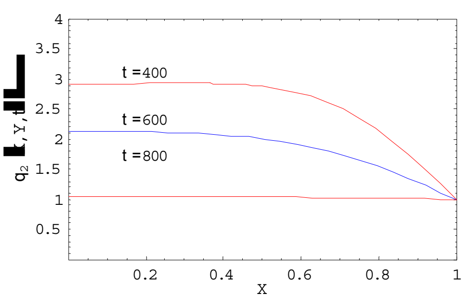


Figure 4. Evolution of moisture profiles during the process

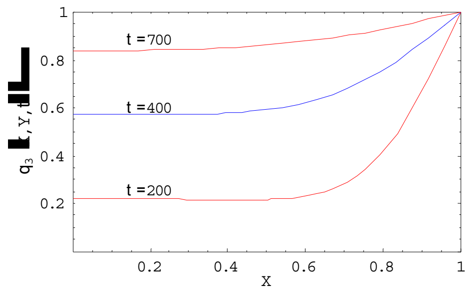


Figure 5. Evolution of pressure profiles during the process

5. Conclusion

In this paper the multidimensional drying problem in a capillary porous media was analytically solved for the associated temperature, moisture and pressure distributions, using Luikov´s model. The generalized integral transform technique (G.I.T.T.) was applied to the problem. Convergence behavior of the adopted numerical methods and results of the temperature, moisture and pressure distributions showed very interesting aspects of drying process on such porous media.

6. References

[1] Luikov, A.V., 1975, "System of Differential Equations of Heat and Mass Transfer in Capillary Porous Bodies (Review)", Int. J. Heat & Mass Transfer, vol. 18, pp. 1- 44

[2] Cotta, R.M., 1993,"Integral Transforms in Computational Heat and Fluid Flow"; CRC Press, Boca Raton, Florida

[3] Cotta, R.M and Mikhailov, M. D., , 1997,"Heat Conduction", John Wiley & Sons.

[4] Duarte, J.B.F., 1995, "Solução Numérica da Difusão de Calor e Massa em Meios Bidimensionais Capilares Porosos", Master Thesis, U.F.C, Ceará, Brasil.

[5] Duarte, J. B. F., 1998, "Abordagem Analítico-Numérica em Meios Multidimensionais Capilares Porosos". Phd Thesis,. Physics Department -UFC. Fortaleza, Ce, Brasil

[6] Ribeiro, J.W., Cotta, R.M. & Mikhailov, M.D., 1993, "Integral Transform Solution of Luikov's Equations for Heat and Mass Transfer in Capillary Porous Media", Int. J. Heat & Mass Transfer, vol. 36, pp. 4467-4475.

[7] Ribeiro, J.W. & Cotta, R.M., 1995, "On the Solution of Non-Linear Drying Problems in Capillary Porous Media Through Integral Transformation of Luikov Equations", Int. J. Num. Meth. Eng., vol. 38, pp. 1001-1020.

[8] Ferguson, W. J., Lewis, R. W. and Tomosy, L., 1993, "A Finite element analysis of freeze-drying of a coffee sample", Computer Methods in Applied Mechanics and Engineering, V. 108, p 341-352.

[9] Lewis, R. W., Morgan, K., Thomas, H. R., and Seetharamu, K. N., 1996, "The Finite Element Method in Heat Transfer Analysis", John Wiley & Sons.

[10] Thomas, H.R., Morgan, K. & Lewis, R.W., 1980, "A Fully Non-linear Analysis of Heat and Mass Transfer Problems in Porous Bodies", Int. J. Num. Meth. Eng., vol. 15, pp.1381-1393.

[11] Mikhailov, M. D. and Özisik, M. N., 1984, "Unified Analysis and Solutions of Heat and Mass Diffusion", John Wiley & Sons, New York.

[12] Duarte, J.B.F. and Ribeiro, J.W., 1997, "Integral Transform Solution of Two-Dimensional Coupled Parabolic Differential Equations, Conferência Italia-Latinoamericana sobre Matemática Aplicada e Industrial - ITLA97", Roma, Itália.

[13] Cunha, F. G. M., Ribeiro, J. W., Duarte, J. B. F., Santos, C. A. C. e Marcondes, F., 2002, "Efeito da Pressão no Problema de Luikov", National Congress of Mechanical Engineering – CONEM, João Pessoa, Paraíba, Brasil.

[14] Wolfram, S., 1996, "The Mathematica Book – Version 3". Ed. Cambridge Univ. Press.

Symbols

a_m	Moisture diffusdion coefficient
c_m	Specific moisture capacity
c_p	Air capacity
c_q	Heat capacity
k_m	Coefficient of moisture conductivity
k_p	Moisture filtration coefficient
k_q	Thermal conductivity
θ_1	Dimensionless temperature distribution
θ_2	Dimensionless moisture distribution
θ_3	Dimensionless Pressure distribution
X	Dimensionless co-ordinate
Y	Dimensionless co-ordinate
τ	Dimensionless time.
δ	Thermo-gradient coefficient
λ	Latent heat
ε	Ratio of vapor diffusion coefficient to the coefficient of total moisture diffusion.

Design of Man Machine Interface for Real-Time Online Control of DC Drives

Jagadeesh Chandra A.P[1], and R.D Sudhaker Samuel[2]
[1]Research Scholar, J.S.S Research Foundation, S.J College of Engineering, Mysore, India
[2]Electronics & Communication Department, S.J College of Engineering, Mysore, India

Abstract - **Automation system developers focus on Internet based control due to availability of several tools to integrate the hardware control equipment with World Wide Web. The paper describes the development of web interface to a DC drive using Virtual Instrumentation tools. The custom electrical hardware serves as the web interface, where it supports various features to control and measure the parameters of the electrical machine remotely. The novel techniques have been used to interface the low power data acquisition system with the DC machine driven by the AC power line supply. The system uses the client-server architecture to access the virtual instruments through the web browser. The developed system allows the remote control and measure of electrical parameters though the Internet.**

I. Introduction

Remote access and control of equipment through the Internet is an interesting research area. Internet based control of a complex hardware is an appealing issue in remote laboratories and industrial applications. The web based applications potentially covers all fields of science and engineering. Developments in computer networking have provided new possibilities and new challenges for designing and deploying remote engineering applications through a commonly available web browser.

The demand for remote engineering applications has been increased due to the advancements in Information Technologies tools which benefits with friendly access, near real-time operation, visual access and efficient utilization of available bandwidth. There are standard technology tools available for remote access of electronic hardware, which can be used for either simple control applications or remote laboratory experiments [4]. The same techniques are used to demonstrate embedded control applications with an additional server to handle the programming of the microcontroller remotely [5]. The client-server architecture is used in these applications with the remote hardware is connected to server. The Hyper Text Mark-up Language (HTML) pages are created for each control applications which are accessed through the Internet. The common feature with them is the operating power range is very small and the complete interface aspects are performed by standard low power data acquisition system connected to web server [7]. The extension of same techniques to electrical machines with the design of novel interface circuitry is described in this paper.

One of the important applications of remote engineering is the development of e-learning methods in engineering with remote laboratories [2]. The complex tasks in electrical engineering cannot be successfully taught without support from new technologies. The Man Machine Interface (MMI) for the DC machine is developed to extend its application to remote experimentation [1]. One of the challenges in remote experimentation is that to give real world experience with control and measurement similar to the manual operation [3]. The system is designed to imitate the manual operation in control functions and various meters are used to record the electrical parameters of the system. The DC machine is driven by the electrical power line supply and the sophisticated interface structure is designed to address the safe and user friendly interface with accuracy in control and measurement operations.

The system architecture is based on client/server model where the communication uses the standard TCP/IP protocol. The standard data acquisition system consisting of DAQ PCI6251 and Electronics Laboratory Virtual Instrumentation Suite (ELVIS) from National Instruments, and the custom designed hardware are used to control various functions of electrical machine. The LabVIEW built-in web server handles the secured access of clients with the required amount of control to the remote device.

II. System Architecture

The proposed architecture is composed of a DC drive connected to the web server by means of a custom designed hardware circuitry and the data acquisition system. The system architecture is shown in Fig. 1.

The architecture is divided in to three sections.

(i) Web-based client-server model with DAQ system
(ii) Custom hardware to control and monitor DC drive
(iii) Input power control and rectification

The web server is the central entity in establishing the communication between the remote clients and the electrical hardware. The web server is primarily used for data acquisition and control operations. LabVIEW with its intuitive graphical

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_44,

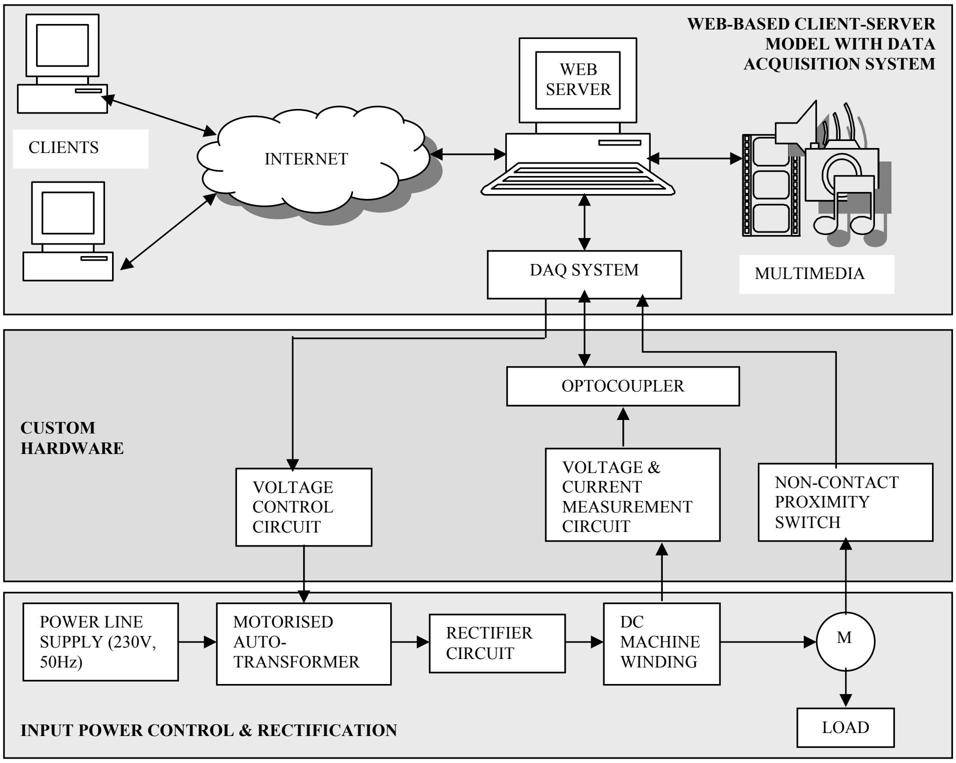


Fig. 1. System Architecture for remote access and control of DC drive.

interface provides the development platform to control the device operation remotely and display the parameters in real-time [6].

The graphical interface to the client side is developed by creating a HTML page of the control panel, which includes the visual display of control commands and meters to measure the electrical parameters of the system. The multimedia features are used to experience the real environment remotely and facilitate the collaborative control between the server and the remote user.

The custom hardware consists of various elements to interface the high power electrical systems with the data acquisition system connected to web server. The input DC voltage to the motor is derived from the AC mains and rectifiers. The system is designed in analogous to manual voltage control, but operated remotely. It is automated using the electrical circuitry which is controlled by the signals from the DAQ system present at the web server. Due to the power constraints of the DAQ PCI6251, all the signal interfaces with the external circuitry has to be limited to 10V. The data acquisition system and the external power circuitry are completely isolated by optocoupler. The non-contact proximity switch is used to measure the speed of the DC motor, where the digital pulses are generated for each rotation of the shaft and counting the number of pulses per second gives the resulting RPM.

The major contributions of this development are listed below.

- The input voltage control variations 0-230V is achieved remotely at the steps of 1V by operating the digital switches using computer keyboard or mouse.
- The measurement range 0-230V, 0-20A is displayed on the remote control panel with perfect isolation between web server and electrical hardware.
- The system is well suited for remote experimentation on DC drives with the adoption of collaborative e-learning tools.

The last section of the system is the well known electrical circuitry used in the laboratory activities. It uses the motorized auto-transformer and rectifier circuitry between the AC mains and the DC drive. The control circuits are designed to interface the DC servo motor of the autotransformer with server, where the digital control commands from server operates the autotransformer. The circuits to measure voltage and current of the motor are directly connected to armature and field windings and are down-converted and fed to web server through the data acquisition system.

The control and measurement of electrical parameters of the remote hardware is achieved through the LabVIEW virtual instruments (VI). The front panel, where the control and measurements are accomplished, is remotely accessible as a web page through the internet.

III. Web interface to electrical hardware

The custom built hardware is the electric circuits used to interface the electrical hardware with the data acquisition system connected to the web server. The data acquisition system consists of NI PCI6251 and ELVIS workstation, which supports 12 channels of Analog Inputs(AI0-AI11) in Referenced Single Ended (RSE) mode, 2 channels of analog output and 8 channels of independent Digital I/O's (Input: DI0-DI7, Output: DO0-DO7). The digital outputs are used to generate control commands to external power hardware and analog inputs are used interface measurement circuitry with virtual instruments. The signals of digital I/O are TTL compatible with 8 bit resolution [8]. Table 1 describes the ELVIS interface with the DC motor control and measurement circuitry. The digital output DO0 controls the main power to the DC Motor. The pair of digital outputs is used to control the input voltage at each windings of the motor.

A. Input Voltage Control Interface

Remote control of input voltage to the DC drives is achieved using the motorized autotransformer with its servo motor is controlled by the digital output commands from the web server. The motorized auto transformer controls the input to the rectifier circuit. The user can set the position of the autotransformer wiper by operating the virtual instruments either through mouse or keyboard, but accurate settings can be made using keyboard control. The basic electrical diagram of the servo control mechanism is shown in Fig. 2.

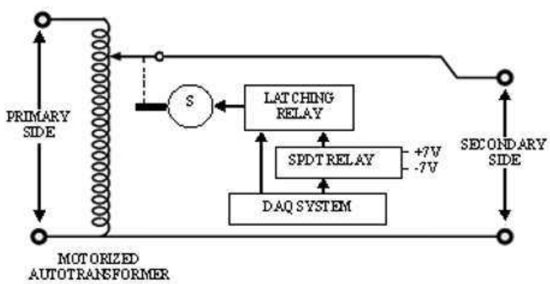


Fig. 2. Input voltage control interface with DAQ system.

TABLE 1
ELVIS I/O PORTS INTERFACE WITH CUSTOM HARDWARE

SI. No.	ELVIS I/O Channel	Function
1	DO0	ON/OFF
2	DO1, DO2	Field Voltage (Control)
3	DO3, DO4	Armature Voltage (Control)
4	AI0	Field Voltage (Measure)
5	AI1	Field Current(Measure)
6	AI2	Armature Voltage (Measure)
7	AI3	Armature Current(Measure)
8	AI4	Speed/RPM (Measure)

The hardware consists of motorized variable output auto transformer, 12V DC servo motor, Single Pole Double Throw (SPDT) relay, latching relay and data acquisition system. The digital control signals from the data acquisition system controls the SPDT and latching relays to operate the servomotor with +/-7Volts. The connected polarity of the DC supply decides the direction of rotation of the motor which in turn moves the autotransformer wiper in either increase or decrease the output voltage. The DC servomotor is operated with +/-7V DC which gives the secondary side voltage control in steps of +/-1V. Hence the output of the autotransformer is digitally controlled for the complete range of 0-250V, 50Hz, AC signal in steps of 1V.

B. Isolation and Down Conversion Logic

This logic is developed to measure the electrical parameters of the DC Drive through LabVIEW virtual instruments. The voltage and current ratings of the custom built electrical circuitry is DC 0-230V, 0-20A respectively. The conversion from 0-230V DC signal to 0-10V DC and interface this signal to data acquisition system with perfect isolation is performed at this stage. The various components constructs this logic includes attenuator, V/F (Voltage to Frequency) converter, optocoupler, F/V (Frequency to Voltage) converter and a low pass filter. The logic used in the conversion process is shown in Fig. 3(a).The V/F and F/V converter provides linear conversion; hence the accuracy in measurement is improved. The low pass filter (LPF) outputs the DC signal without ripples present in it.

The current measurement follows the same principle with minor changes in the signal path. The conversion stages for current measurement are shown in Fig. 3(b).

Initially the current in the motor windings are converted to equivalent voltage by connecting a small resistor in the current path and the voltage across it is taken for the measurement. The small voltage proportional to current is suitably amplified and fed to the input channel of DAQ system. The same method used for voltage measurement is followed with a buffer at the final stage.

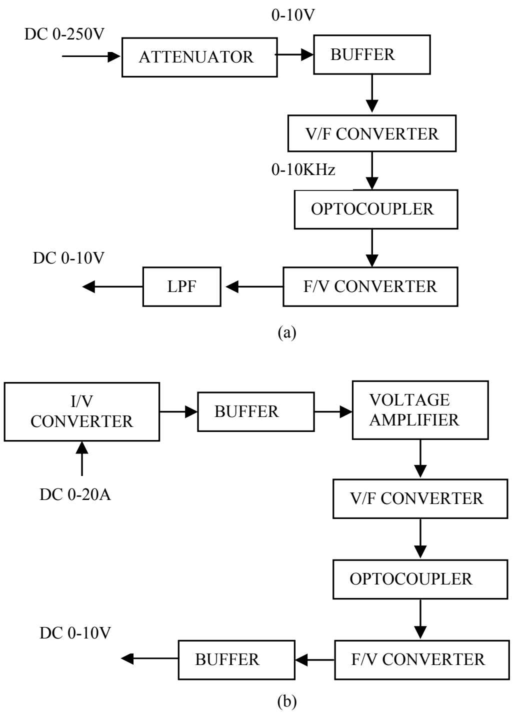


Fig. 3. Block diagram of Isolation and Down-conversion logic.
(a) Voltage down-conversion (b) Current down-conversion

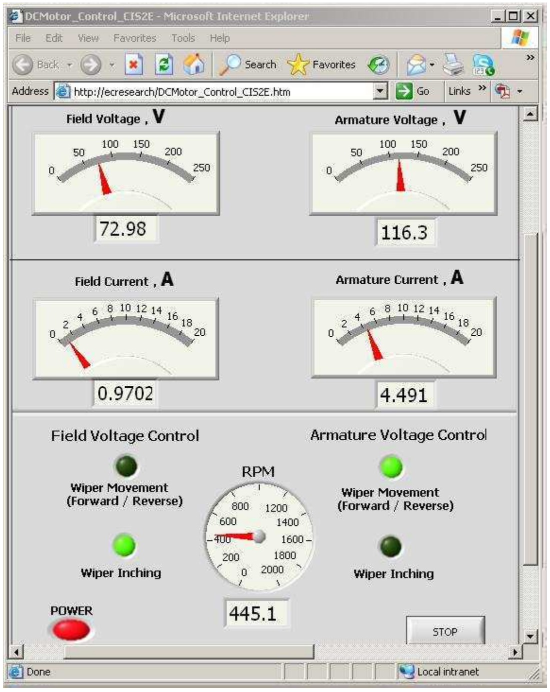


Fig. 4. Remote Front Panel for Control and Measurement

IV. Remote Control Panel

The graphical programming language LabVIEW is used in the development of GUI (Graphical User Interface) and it offers several methods for remote access using TCP/IP protocol without additional programming. The LabVIEW VIs developed to control and measure the parameters of the remote hardware is shown in Fig. 4. The LabVIEW programs are called Virtual Instruments (VI) because their appearance and easy operation imitate real instruments. A VI mainly contains a "front panel" which serves as the user interface and a "block diagram" contains the graphical code. The LabVIEW web server is used to create and publish HTML documents of the front panel on the web and embed VIs in a web page. The VI front panel can be viewed remotely, either from within LabVIEW or from within a web browser, by connecting to the LabVIEW built-in web server. When the front panel remotely opened through the web, the web server sends the front panel to the client, but the block diagram and all other sub VIs remain on the server computer. The remote client can interact with the front panel in the same way as if the VI were running on the client, except the block diagram executes on the server [9]. This feature is used to publish entire front panel or to control the remote application safely and easily. The meters read the actual values after the scaling operation are performed through the graphical code The client side uses the "LabVIEW run-time engine", freely downloadable software to execute the commands in remote environment.

V. Conclusion

On-line control of DC machine for remote laboratory application is described in this paper. The integration of advanced tools with existing facilities enhances the flexibility and ease in operation without much additional costs. The DC machine is driven by the electrical power line supply and the sophisticated interface structure is designed to address the safe and user friendly interface with accuracy in control and measurement. The flexible and versatile design and implementation of the system is suitable to be used in the areas where remote interaction with the electrical drives is required. This specific design is intended to be used in the remote laboratory applications on DC drives.

ACKNOWLEDGEMENT

This paper reports the partial results of a major research on Remote Access Labs funded by the Department of Science and Technology, Ministry of Science and Technology, Government of India, Vide Number: SR/FST/ETI-033/2005.

REFERENCES

[1] C. Guerra Torres, J. de Leon Morales, A. Glumineau, D. Traore, R. Boisliveau. "Teleoperation of an experimental platform of electrical machines through the internet", International Journal of Online Engineering (iJOE), 4(1), 2008.

[2] Emi Ernest, Rafal Sztylka, Bartlomiej Ufnalski, Wlodzimierz Koczara, "Methods in teaching modern ac drives: inverter-fed motor system with internet-based remote control panel", Proceedings of 12th International Power Electronics and Motion Control Conference, Portoroz, Solvenia, 2006, pp.2130-2133.

[3] Viliam Fedak, Pavol Bauer. "E-learning in education of electrical drives and power electronics: opportunities and challenges", International Conference on Electrical Drives and Power Electronics (EDPE), Dubrovnik, Croatia, 2005.

[4] Shuhui Li, Abrar A. Khan, "Applying it tools to a laboratory course for measurement, analysis, and design of electric and electronic circuits" IEEE Transactions on Education, Vol.48, No.3, pp.520- 530, 2005.

[5] M.Gilibert, J.Picazo, M.Auer, A.Pester, J.Cusido, J.A. Ortega. "80C537 Microcontroller remote lab for e-learning teaching", iJOE International Journal of Online Engineering, vol.2, 2006

[6] M.B. Naumovic, D. Zivanovic. "Remote experiments in control engineering education laboratory", iJOE International Journal of Online Engineering, 4(2), 2008, pp.48-53.

[7] William Gerard Hurley, Chi Kwan Lee. "Development, implementation, and assessment of a web-based power electronics laboratory", IEEE Transactions on Education, 48(4), 2005, pp.567-573.

[8] NI Educational Laboratory Virtual Instrumentation Suite (NI ELVIS) user manual, National Instruments, National Instruments Corporation, 2006.

[9] LabVIEW 7 Express user manuel, National Instruments, National Instruments Corporation, Ireland, 2003.

The Output Regulation for Tracking Unusual Motion Patterns by Active Disturbance Rejection Control

Wei Wang
Information School, Renmin University of China,
No. 59, Zhongguancun, Beijing 100872, P. R. China

Abstract *–In this paper, from the viewpoint of engineering applications, we consider the tracking regulation for unusual motion pattern such as sawtooth, square wave etc. The method can be summarized as follows. Firstly, we use a dynamic mechanism to regulate the gain of Integral part of PID controller such that the system can eliminate the disturbances which are form the system itself or exogenous, and then the system can be synthesized easily. Secondly, we design the control input through proper forms of nonlinear combination of Proportional and Derivative of error for the remainder system which without disturbances. The method can reject uncertainties or disturbances effectively and implement output regulation powerfully even though the system contains uncertainties or suffers form disturbances, and even the targets' movement is unusual. The method can also be extended to higher order systems or to MIMO interconnected systems. The theoretical analysis and simulation results are given to indicate the effectiveness of the method.*

Keywords *–unusual motion, output control, PID control, ADRC*

I. INTRODUCTON

It is well known that PID control is still widely used in process controls because of its robustness and with a simple structure ([1]--[3]). Modern control theories in recent decades don't changed the fundamental status of PID control in real-time process controls ([4], [5]).

Based on many years development of PID control ([3], [6]), especially nonlinear PID control ([6], [7]), another kind of control method ---ADRC was proposed ([8]). The method has gotten developed in recent years ([8]--[10]). It can not only get rid of the influence of disturbances but also get better control performance as well. The basic principle of ADRC is that it uses the extended state observer (ESO) ([7]) to estimate the influence of disturbances or uncertainties indirectly, and force the system changing in a normalized way. And then it constructs a new control input for the remainder system which without disturbances. Because of its strong ability to reject disturbances, it has been expanded widely (see [8]--[10] and references therein).

In this paper, we consider the output regulation for tracking unusual motion patterns. We use a dynamic process to adjust the gain of Integral part of PID control to eliminate the uncertainties or disturbances. And then by using proper nonlinear combinations we assign the structure of control input for the remainder system which without disturbances. We also consider the cases such as higher order systems and MIMO interconnected systems, etc. The main idea and theoretical results are provided. Some simulation results are shown to indicate the effectiveness of the method.

The benefits of the control method are as follows. It can get rid of the influences of disturbances or uncertainties effectively and make the systems' synthesis easily. Based on the using of tracking-differentiators, it can make a proper tracking for targets which move in the way of undifferential, or even discontinuous. It can also extend the functions of ordinary PID controllers to higher-order of systems. The most important thing of all is that the method is more a kind of control technique than a kind of control theory. Therefore it is easy for engineering applications. It is the new form of ADRC that, without estimating the influence of disturbances, we can still get rid of the impact of disturbances effectively, and can fulfill powerful output regulation as well. Theoretical analysis and simulation results indicate the efficiency of the method. From the viewpoint of control theory, it can also implement the linearization and decoupling even for uncertain systems, which is impossible form the viewpoint of differential geometry method.

The paper is organized as follows. In section 2, the basic idea of the method, which we call it the new form of ADRC, is described. Section 3 provides the main results about the new form of ADRC. In section 4, some simulation results are shown to indicate the efficiency of the method.

II. THE STATEMENT OF PROBLEM

To implement output tracking, the practical method is PID control. Suppose that $y_r(t)$ is the reference input, $y(t)$ is the output of system. The error is $e(t) = y_r(t) - y(t)$, then the conventional PID control input is:

$$u(t) = a_0 \int_{t_0}^{t} e(\tau)\, d\tau + a_1 e(t) + a_2 \dot{e}(t) \tag{1}$$

where a_0, a_1, a_2 are the design parameters of integral, proportional and derivative respectively.

To avoid the contradiction of overshoot and fastness in output response process, the nonlinear PID controller was provided ([6]), and the control input can be chosen as follows:

$$\begin{aligned} u(t) = a_0 \left| \int_{t_0}^{t} e(\tau) d\tau \right|^{\alpha} \operatorname{sgn}\left(\int_{t_0}^{t} e(\tau) d\tau \right) + \\ + a_1 |e(t)|^{\alpha} \operatorname{sgn}(e(t)) + a_2 |\dot{e}(t)|^{\alpha} \operatorname{sgn}(\dot{e}(t)) \end{aligned} \tag{2}$$

where $0 < \alpha \le 1$ is also a design parameter.

One of the most important problems in PID control is the tuning of the parameters a_0, a_1, a_2 in (1) or (2). It is still be

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_45,

considered in many papers ([3] and [9]). In author's opinion, the main difficulties in tuning the parameters of PID controllers are the following two aspects. One is the uncertainties or disturbances encountered. The other is that all parameters should be tuned simultaneously.

Now, we will turn the problem of parameters' tuning into disturbances' rejection plus the control system synthesis for the remainder system which contains no disturbances. The main idea can be summarized as follows.

For the sake of simplicity, we take the following form of error system for example

$$\begin{cases}\dot{e}_1 = e_2 \\ \dot{e}_2 = f(e_1, e_2) + d(t) + u\end{cases} \quad (3)$$

where $d(t)$ is the disturbance which comes from the internal or exogenous. Here, the disturbance means, in addition to the one as usual, the unmodeled part, or even the generalized derivative which is gotten by using tracking-differentiators for preprocessing of the unusual tracking targets. Therefore, in order to solve the control problem, based on the elicitation of the variable structure control method ([4]), we hope to eliminate or reject the disturbance first, and then force the system (3) changes in following way:

$$\begin{cases}\dot{e}_1 = e_2 \\ \dot{e}_2 = f(e_1, e_2) + \bar{u}\end{cases} \quad (4)$$

How to meet the need stated above? We will prove that disturbances can be rejected by introducing dynamic mechanism into PID controllers to adjust the gain of the integral part of the PID controller. And then we can use the proper combination of Proportional and Derivative as the control input ($\bar{u}$) for the system without disturbances. That is the basic idea of the new form of ADRC we provided here.

If the disturbances can be taken away from the system, the new control input $\bar{u}$ can be chosen in the form of linear or nonlinear one which depends on the form of $f(e_1, e_2)$.

In order to improve the efficiency of output regulation for tracking unusual targets, based on the idea of nonlinear PID controllers ([6]), the control input $\bar{u}$ can be chosen as:

$$\bar{u}(t) = a_1 |e(t)|^{\alpha} \operatorname{sgn}(e(t)) + a_2 |\dot{e}(t)|^{\alpha} \operatorname{sgn}(\dot{e}(t)) \quad (5)$$

where $0 < \alpha \le 1$ is a given constant.

So the main problem here is how to accomplish the goal stated above. And if it is possible, after rejecting the disturbances, the remaining problem is to assign the control input $\bar{u}$ or to choose its parameters.

III. THE MAIN RESULTS

The key to the problem here is the choice of proper dynamic mechanism which used for regulating the gain of Integral part of PID control. Here, we choose the following form which is the same as given in the book ([5]). That is to say, we introduce the dynamic process $\mu(t)$ described by the following differential equation:

$$\dot{\mu}(t) = \begin{cases}-\gamma \operatorname{sgn}(\sigma(e)), & as\ |\mu(t)| \le 1 \\ -\omega\mu, & as\ |\mu(t)| > 1, \quad \mu(0) = \operatorname{sgn}(\sigma(e(0)))\end{cases} \quad (6)$$

where $\sigma(e)$ is the functional of control error, it can be determined by control requirements, and $\omega > 0$ is a given positive constant. Therefore the control input is:

$$u(t) = a_0 \mu \int_{t_0}^{t} |e(\tau)|\, d\tau + \bar{u}(t) \quad (7)$$

where $\bar{u}$ is the control input for the remainder system which contains no disturbances.

Next, we will pay attention to the possibility of choosing the design parameters γ and a_0. We consider the system of (3), (6) and (7). Obviously, it is described by differential equations with discontinuous righthand sides. The existence and extension of its solution can be considered according to [11].

Lemma 1. For the solution $\mu(t)$ of equation (6), in which $\omega > 0$ is a given constant, it is evident that

$$|\mu(t)| \le 1 \quad (8)$$

And if $\sigma(e)$ keeps sign in interval $[t_1, t_2]$, and $t_2 - t_1 > \frac{2}{\gamma}$, it can be shown easily that

$$\mu(t_2) = -\operatorname{sgn}(\sigma(e(t_2))) \quad (9)$$

Based on the basic characteristics of $\mu(t)$, we will get related results about the design parameters of γ and a_0 for certain different cases.

A. The basic estimation of design parameters for second order uncertain systems.

We will consider the estimation of design parameters for second-order systems first.

Theorem 1. Consider the second-order uncertain system

$$\begin{cases}\dot{e}_1 = e_2 \\ \dot{e}_2 = f(e_1, e_2) + d(t) + a_0 \mu \int_{t_0}^{t} |e_1(\tau)|\, d\tau + \bar{u}(t)\end{cases} \quad (10)$$

where $d(t)$ is uncertain part, $\mu(t)$ is given by (6), and let

$$\sigma(e) = d(t) + a_0 \mu \int_{t_0}^{t} |e_1(\tau)|\, d\tau \quad (11)$$

Then as

$$a_0 > \sup_{t > t_0} \left| \frac{d(t)}{\int_{t_0}^{t} |e_1|\, d\tau} \right| \quad (12)$$

and

$$\gamma > \sup_{t > t_0} \left| \frac{a_0 |e_1| + \frac{d}{dt}[d(t)]}{a_0 \int_{t_0}^{t} |e_1|\, d\tau} \right| \quad (13)$$

hold, there is a finite time t', as $t \ge t'$, the following equality holds

$$d(t) + a_0 \mu \int_{t_0}^{t} |e_1(\tau)|\, d\tau = 0 \quad (14)$$

Proof: Getting the derivatives for both sides of (11), we have

$$\dot{\sigma}(e) = \dot{d}(t) + a_0\dot{\mu}\int_{t_0}^{t} | e_1(\tau) | \, d\tau + a_0\mu|e_1(\tau)|$$

It should be mentioned that the derivative of $d(t)$ may be a generalized one when $d(t)$ is a undifferential function.

According to (8) we know $|\mu(t)| \le 1$, then

$$\dot{\mu}(t) = -\operatorname{sgn}(\sigma(e_1(t)))$$

So, if (13) is true, we have the following inequality.

$$\sigma\dot{\sigma} < 0$$

Next, we will prove that there is a finite time $t_1^{'}$ such that $\sigma(t_1^{'}) = 0$.

As a matter of fact, if σ is to be nonzero form t_1, i.e. σ keeps sign, then according to Lemma 1, as $t > t_1 + \frac{2}{\gamma}$, we have the following equality

$$\mu(t) = -\operatorname{sgn}(\sigma(e_1(t))) \tag{15}$$

Therefore from (12) and (15) we know that σ has the sign opposite to itself. So there is a contradiction, i.e., there is a finite time $t_1^{'}$ such that $\sigma(t_1^{'}) = 0$.

According to the analysis above we know that Theorem 1 holds.

From Theorem 1 we know that, under the given conditions, we can choose γ and a_0 to reject disturbance directly. So we needn't known the specific form of $d(t)$, and we need only know the upper and lower variant boundary of $d(t)$ and $\dot{d}(t)$. And the other parameters can be assigned by the requirements of performance which should keep stability of the remainder system. From the viewpoint of parameters' tuning in PID control, the method can turn the conventional tuning method which should choice the three parameters simultaneously into the method that the parameters can be determined separately.

On the other hand, for engineering implementation, because of $d(t)$ is often unknown to us, we can choose $\sigma(e)$ in the following form:

$$\sigma(e) = \dot{e}_2 - f(e_1, e_2) - \bar{u}$$

In such a case, we can avoid the difficulty of getting $d(t)$ and $\dot{d}(t)$ directly.

B. Results for nth-order SISO systems.

In this subsection we extend above results to nth-order SISO systems. In such cases, in order to improve the ability to reject disturbances, we make a slight change to above results, i.e., we add a constant *k* or parameter to the Integral part of PID controllers. Similar to the proof of Theorem 1, we have the following results.

Theorem 2. Consider the nth-order SISO system

$$\begin{cases} \dot{e}_1 = e_2 \\ \dot{e}_2 = e_3 \\ \cdots \\ \dot{e}_n = f(e_1, \cdots, e_n) + d(t) + a_0\mu\left(\int_{t_0}^{t} | e_1(\tau) | \, d\tau + k\right) + \bar{u} \end{cases} \tag{16}$$

where $\mu(t)$ is decided by (6) and

$$\sigma(e) = d(t) + a_0\mu\left(\int_{t_0}^{t} | e_1(\tau) | \, d\tau + k\right) \tag{17}$$

Then as

$$a_0 > \sup_{t > t_0} \left| \frac{d(t)}{\int_{t_0}^{t} | e_1 | \, d\tau + k} \right| \tag{18}$$

and

$$\gamma > \sup_{t > t_0} \left| \frac{a_0 \, | e_1 | + \frac{d}{dt}[d(t)]}{a_0 \int_{t_0}^{t} | e_1 | \, d\tau + k} \right| \tag{19}$$

hold, there is a finite time $t'(\ge t_0)$ such that, as $t \ge t'$, the following equality holds

$$d(t) + a_0\mu\left(\int_{t_0}^{t} | e_1(\tau) | \, d\tau + k\right) = 0$$

And the result also indicates that we can extend the former PID control method form second-order systems to higher-order systems.

C. The analysis results for interconnected systems.

For interconnected systems, we can regard the interconnected terms as disturbances. Therefore the above results can be extended to following systems.

Theorem 3. Consider the following interconnected system

$$\begin{cases} \dot{e}_{i1} = e_{i2} \\ \dot{e}_{i2} = f(e_{i1}, e_{i2}) + d_i(t) + a_{i0}\mu_i \int_{t_0}^{t} (| e_{11} | + \cdots + | e_{m1} |) d\tau \\ \quad + g_i(e_1, \cdots, e_m) + \bar{u}_i(t) \end{cases} \tag{20}$$

where $i = 1, \cdots, m$, $g_i(e_1, \cdots, e_m)$, $d_i(t)$ are the interconnected terms, and $\mu_i(t)$ ($i = 1, \cdots, m$) are given by (6), and let

$$\begin{aligned} \sigma(e) = {} & a_{i0}\mu_i \int_{t_0}^{t} (| e_{11} | + \cdots + | e_{m1} |) d\tau \\ & d_i(t) + g_i(e_1, \cdots, e_m) \end{aligned} \tag{21}$$

Then as

$$a_{i0} > \sup_{t > t_0} \left| \frac{d_i(t) + g_i(e_1, \cdots, e_m)}{\int_{t_0}^{t} | e_{11} | + \cdots + | e_{m1} d\tau} \right| , \ i = 1, \cdots, m \tag{22}$$

and

$$\gamma > \sup_{t>t_0}\left|\frac{a_{i0}(|e_{11}|+\cdots+|e_{m1}|)+\frac{d}{dt}[d_i(t)]}{a_{i0}\int_{t_0}^{t}(|e_{11}|+\cdots+|e_{m1}|)d\tau}\right|, \ i=1,\cdots,m \quad (23)$$

hold, there is a finite time $t'(\geq t_0)$, as $t \geq t'$, the following equality holds , for $i=1,\cdots,m$

$$a_{i0}\mu_i\int_{t_0}^{t}(|e_{11}|+\cdots+|e_{m1}|)d\tau + d_i(t) + g_i(e_1,\cdots,e_m) = 0 \quad (24)$$

Proof: Similar to the proof of Theorem 1 we know that Theorem 3 holds.

According to the conclusion above we know that, through proper choice of the design parameters, we can also realize decoupling of the interconnected systems by the method given in this paper.

Similar to above analysis we can get results for more general form of MIMO systems.

Theorem 4. Consider the nth-order SISO system

$$\begin{cases}\dot{e}_{i1} = e_{i2}\\ \dot{e}_{i2} = e_{i3}\\ \cdots\\ \dot{e}_{ir_i} = f_i(e_{i1},\cdots,e_{ir_i}) + d_i(t) + a_{i0}\mu_i\left(\int_{t_0}^{t}\sum_{i=1}^{m}|e_{i1}(\tau)|d\tau + k_i\right) +\\ \quad + g_i(e_1,\cdots,e_m) + \bar{u}_i\end{cases} \quad (25)$$

where $i=1,\cdots,m$, $g_i(e_1,\cdots,e_m)$, $d_i(t)$ are the interconnected terms, and $\mu_i(t)$, $i=1,\cdots,m$ is given by (6), and let

$$\sigma(e) = a_{i0}\mu_i\int_{t_0}^{t}(|e_{11}|+\cdots+|e_{m1}|)d\tau + d_i(t) + g_i(e_1,\cdots,e_m) \quad (26)$$

Then as

$$a_{i0} > \sup_{t>t_0}\left|\frac{d_i(t)+g_i(e_1,\cdots,e_m)}{\int_{t_0}^{t}|e_{11}|+\cdots+|e_{m1}d\tau}\right|, \ i=1,\cdots,m \quad (27)$$

and

$$\gamma > \sup_{t>t_0}\left|\frac{a_{i0}(|e_{11}|+\cdots+|e_{m1}|)+\frac{d}{dt}[d_i(t)]}{a_{i0}\int_{t_0}^{t}(|e_{11}|+\cdots+|e_{m1}|)d\tau}\right|, \ i=1,\cdots,m \quad (28)$$

hold, there is a finite time $t'(\geq t_0)$, as $t \geq t'$, the following equality hold, i.e., for $i=1,\cdots,m$

$$a_{i0}\mu_i\int_{t_0}^{t}(|e_{11}|+\cdots+|e_{m1}|)d\tau + d_i(t) + g_i(e_1,\cdots,e_m) = 0 \quad (29)$$

All these results show that, by the method, we can not only reject disturbances but also realize decoupling for interconnected systems effectively. And then we can design control input for systems which are the ones without disturbance or are decoupling. So the design can be accomplished easily according to classical or modern control methods. At the same time, the method avoids the problem of constructing ESO, which may be difficulty for some systems with complex structures or interrupted by disturbances seriously.

IV. SIMULAITON RESULTS

We present some simulation results for the method discussed above in this section. For a nonlinear uncertain system

$$\begin{cases}\dot{x} = F(x) + Dd(t) + Bu\\ y = C(x)\end{cases} \quad (30)$$

where $x \in R^n$, B, C, D are proper matrices which keep the equation meaningful. y is the output, and assume that it can be measured. $d(t)$ refers to disturbance which is unknown to us. In order to make y follows the given signal y_r which is unusual, we adopt the new form of ADRC for nth-order systems (where $n \geq 2$).

In doing simulation, the tracked signal y_r may be undifferential. Therefore we can preprocess the signal by using tracking-differentiators (TD) ([12]) which is a nonlinear system as follows:

$$\begin{cases}\dot{x}_{11} = x_{12}\\ \dot{x}_{12} = -Rsat\left(x_{11} - y_r(t) + \frac{|x_{12}|x_{12}}{2R}, \delta\right)\end{cases} \quad (31)$$

From [12] we know that, as $R \to \infty$, we will have the results such as $x_{11} \to y_r$ and $x_{12} \to \dot{y}_r$. So we can regard x_{12} as the (generalized) derivative of y_r .

A. Getting higher-order derivatives

First, after the preprocessing of y_r and y , we can get the error and its higher-order derivatives:

$$\begin{cases}e_0 = \int_{t_0}^{t}(x_{11} - x_{21})d\tau\\ e_1 = x_{11} - x_{21}\\ e_2 = x_{12} - x_{22}\\ e_3 = x_{14} - x_{24}\end{cases} \quad (32)$$

where e_1 is the error, e_2 is the first-order derivative of error, and e_3 is the second-order derivative.

Then, for second-order system, let $\mu(t)$ determined by

$$\dot{\mu}(t) = \begin{cases}-\gamma\,\mathrm{sgn}(\dot{e}_3 + a_1|e_1|^{\alpha}\mathrm{sgn}(e_1) + a_2|e_2|^{\alpha}\mathrm{sgn}(e_2)), & as\ |\mu(t)| \leq 1\\ -\omega\mu, \quad as\ |\mu(t)| > 1, & \mu(0) = \mathrm{sgn}(\sigma(e(0)))\end{cases}$$

And the control input is the following one:

$$\bar{u}(t) = a_0\mu|e(t)|^{\alpha} + a_1|e(t)|^{\alpha}\mathrm{sgn}(e(t)) + a_2|\dot{e}(t)|^{\alpha}\mathrm{sgn}(\dot{e}(t)) \quad (33)$$

where a_0 can be determined by the range of disturbance and disturbance's derivative. And a_1, a_2 can be gotten by assigning which should keep the stability of the following error system:

$$\begin{cases}\dot{e}_1 = e_2\\ \dot{e}_2 = f(e_1, e_2) + a_1|e(t)|^{\alpha}\mathrm{sgn}(e(t)) + a_2|\dot{e}(t)|^{\alpha}\mathrm{sgn}(\dot{e}(t))\end{cases} \quad (34)$$

So we can choose a_0 according to (12), and then assign a_1, a_2 properly. As for other kind of systems, we can deal with them in a similar way.

C. Simulation examples.

In all simulations, let $\omega = 0.5$ in (6). Other parameters can be chosen based on certain given situations.

Example 1. Consider the following second-order nonlinear system

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_1^2 + 1.5x_2^2 + 4.9\sin(t) + u \\ y = x_1 \end{cases} \tag{35}$$

where 4.9sin(t) can be regarded as a term of disturbance. We hope that the output $y = x_1$ can follow the reference signal y_r which takes the form of square and sawtooth waves respectively. Owing to the undifferential of the reference signals, we can use TD to smooth y_r. Here we choose the discrete form of TD ([13]) for precision. The parameters in TD are chosen as r=30, h=0.01, T=0.01. The parameters in PID are given by a_0=20, a_1=90, a_2=90, and the parameter of γ=10. The results on output tracking of the two cases are shown in Fig. 1 and Fig. 2 respectively.

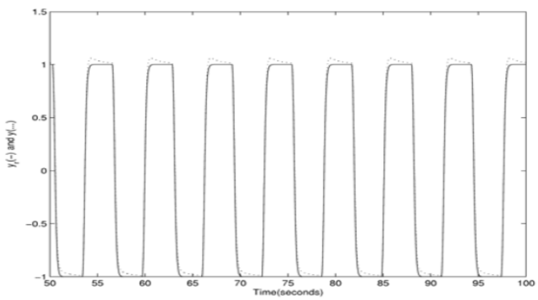


Fig.1 The output result of tracking square wave

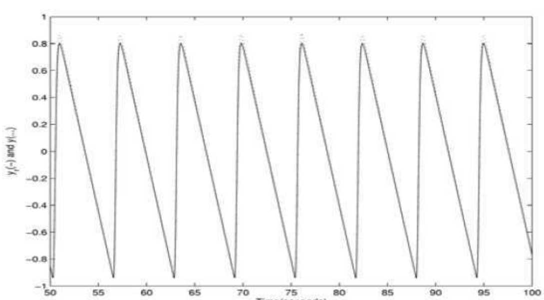
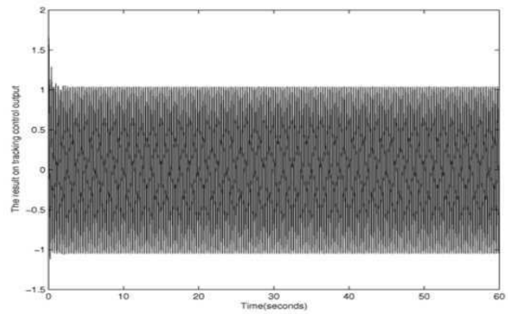


Fig.2 The output result of tracking sawtooth wave

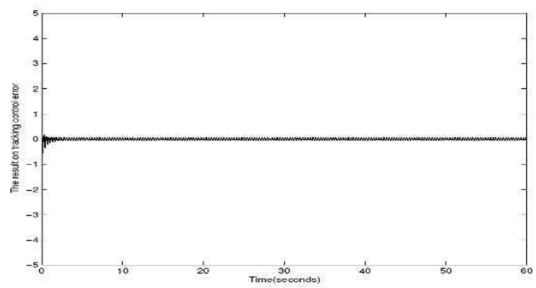
Example 2. Consider the third-order nonlinear system

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 \\ \dot{x}_3 = x_1^2 + x_2 x_3 + \sin(10.5t) + 5.8 + u \end{cases} \tag{36}$$

where sin(10.5t)+5.8 be regarded as a term of disturbance. We hope that the output $y = x_1$ can follow the reference signal y_r. Here $y_r = \sin(20t)$ is the signal with a high frequency. Because the setting signal y_r and the output signal $y = x_1$ are all continuous functions, no TD is used in this simulation. The parameters k, a_0 in (18) and γ in (19) are 100, 0.06 and 10 respectively. The remainder part of control input, that is, the $\bar{u}$ is chosen in the form of $\bar{u}(t) = a_1\mu|e_1(t)|^{\alpha} + a_2|e_2(t)|^{\alpha}\operatorname{sgn}(e_2(t)) + a_3|\dot{e}_3(t)|^{\alpha}\operatorname{sgn}(\dot{e}_3(t))$. And a_1 =18660, a_2 =13678, a_3=50. The results are shown in Fig. 3.



(a) The result on control output of tracking sin(20t)



(b) The result on tracking control error

Fig.3 The results of tracking high frequency function and the tracking error

Example 3. Consider a nonlinear system related with the vehicle dynamics ([14])

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 \\ \dot{x}_3 = a(x_2, x_3) + b(x_2)u \end{cases} \tag{37}$$

where

$$a(x_2, x_3) := -2\frac{K_d}{m}x_2 x_3 - \frac{1}{\tau(x_2)}\left(x_3 + \frac{K_d}{m}x_2^2 + \frac{d_m}{m}\right) \tag{38}$$

and

$$b(x_2) := \frac{1}{m\tau(x_2)} \tag{39}$$

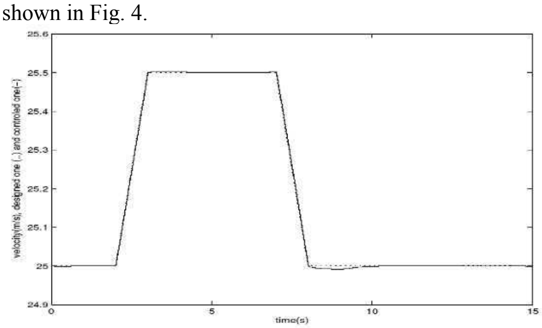
where m is the mass of vehicle, τ is the time constant of its engine, K_d is the aerodynamic drag coefficient, d_m is the mechanical drag, and u is the engine input.

We hope that the trajectory of the vehicle x_1 has a certain given form as follows:

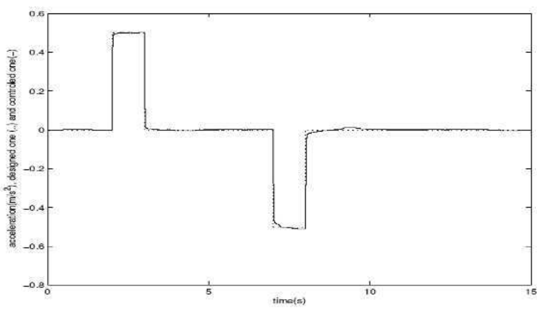
$$
y_r(t) = \begin{cases} 25t, & 0 \le t \le 2 \\ (t-2)^2/4 + 25(t-2) + 50, & 2 < t \le 3 \\ 25.5(t-3) + 75.25, & 3 < t \le 7 \\ -(t-7)^2/4 + 25.5(t-7) + 177.25, & 7 < t \le 8 \\ 202.5 + 25(t-8), & t > 8 \end{cases} \tag{40}
$$

And x_2, x_3 can track $\dot{y}_r$, $\ddot{y}_r$ respectively.

We choose the parameters as follows: in (6) $\gamma = 0.5$, $\omega = 9$; in TD, $r = 30$, $h = 0.01$, $T = 0.01$. $\bar{u}$ is chosen in the form of $v = \sum_{i=1}^{3} a_i |e_i|^{0.6} \operatorname{sgn}(e_i)$, and a_1=19050, a_2=9398, a_3=6731. The tracking results of velocity and acceleration, i.e., x_2, x_3 are shown in Fig. 4.



(a) The result of x_2 tracking $\dot{y}_r$



(b) The result of x_3 tracking $\ddot{y}_r$

Fig.4 The results of x_2, x_3 tracking the given functions

From these results, it can be seen that, using the new form of ADRC, we can get better tracking results not only for unusual signals but also for tracking high frequency signals. At the same time, we can get better results for the systems which with higher orders (see Example 2), and even for the disturbances which are with the 'stochastic' characteristics.

It seems that the new forms of ADRC have more parameters to be tuned than the former one, or the structures are more complex than original PID controllers. As a matter of fact, the main parameters in the new ADRC are a_0 and γ, which can be determined by the theorems given above. The other parameters such as a_1 and a_2 can be chosen by assigning. And the choice of parameters in TDs can be founded in [13]. So, it is easier here than former one. It should be mentioned that the choice of those parameters can be fulfilled separately.

Of course, the main problem here is that we use the integration about the absolute value of error in (7). It may cause amplification of error on disturbances' rejection. This can be avoided by using $\left|\int_{t_0}^{t} e(\tau) d\tau\right|$ instead of $\int_{t_0}^{t} |e(\tau)| d\tau$. But this will not have any influence on the simulation results of output tracking nearly.

V. CONCLUSIONS

It is the method given this paper that the synthesis of control system can be changed into the estimating the boundary of disturbances (instead of the construction of ESO in former ADRC) plus the assignment of other parameters or proper forms of errors for the remainder system which without disturbances. The assignment of the remainder parts can be implemented by the performance of tracking control. And simulation results indicate the effectiveness of the method.

At the same time, from simulation results for above systems, we can get following conclusions.

For a group of given design parameters, the tracking results may keep unchanged nearly even those systems' forms changed in proper ranges. That is to say, the method has certain robustness. To keep the effectiveness of tracking, the design parameters can also be changed in certain ranges. That is to say, the control method also has certain adaptability.

FREFERENCES

[1] K.J. Åström and T. Hagglund, "Automatic tuning of PID controllers," ISA, Research Triangle Park, NC, USA, 1988.

[2] A. Datta, M.T. Ho and S.P. Bhattachoryya, *Structure and Synthesis of PID Controllers,* London: Springer, 2000.

[3] T. Nusret and D. P. Atherton, "Design of stabilizing PI and PID controllers," *Int. J. Systems Science*, vol.37, pp. 543-554. August, 2006.

[4] M. Ö. Efe, C. Ünsal, O. Kaynak and Xinghuo Yu, "Variable structure control of a class of uncertain systems," *Automatica*, vol.40, 59-64, January 2004.

[5] S.V. Emelyanov and S. K. Korovin, *Control of Complex and Uncertain Systems: New Types of Feedback*, Berlin: Springer, 2000.

[6] J.Q. Han, "Nonlinear PID Controllers," *Acta Automatica Sinica*, vol.20, pp.487--490, 1994. (In Chinese).

[7] J.Q. Han, "Extended state observer for uncertain systems," *Control and Decision*, vol.10, pp. 85-88, January 1995, (In Chinese)

[8] J.Q. Han, "Active disturbance rejection control and its application," *Control and Decision*, vol.13, pp.19-23, January 1998, (In Chinese).

[9] J.Q. Han, "Active disturbance rejection control technology," *Frontier Sciences*, vol.1, 2007, (in Chinese).

[10] Y. Xia, P. Shi, G.P. Liu, D. Rees and J. Han, "Active disturbance rejection control for uncertain multivariable systems with time-delay," *IET Control Theory and Applications*, vol.1, pp.75-81, Jan. 2007.

[11] A.F. Filippov, *Differential Equations with Discontinuous Righthand Sides*, Kluwer Academic Publishers, Dordrecht, 1988.

[12] J.Q. Han and W. Wang, "Nonlinear Tracking-Differentiators," J. *Systems Science and Mathematical Sciences*, vol. 14, pp. 177--183, April 1994. (In Chinese).

[13] J.Q. Han and L. Yuan, "The discrete form of Tracking-Differentiators," *J. Systems Science and Mathematical Science*}, vol.19, pp. 268--273, 1999. (In Chinese).

[14] G. Burgio and P. Zegelaar, "Integrated vehicle control using steering and brakes," *Int. J. Control*, vol.79, pp. 534-541, May 2006.

Design of a High Performance, Low Power, Fully Differential Telescopic Cascode Amplifier using Common-Mode Feedback Circuit

Tushar Uttarwar[1] Sanket Jain[2] and Anu Gupta[3]

[1]*Student, Birla Institute of Technology & Sciences, Pilani, tusharuttarwar@gmail.com*
[2]*Student, Birla Institute of Technology & Sciences, Pilani, sanket.jain@ymail.com*
[3]*Assistant Professor, Birla Institute of Technology & Sciences, Pilani, anug@bits-pilani.ac.in*

***Abstract*—This paper describes the design of a high performance, low power fully differential telescopic amplifier. As is well known, high gain differential amplifiers require an additional Common-Mode Feedback (CMFB) circuit owing to their high output impedances. This is because high output impedance makes it difficult to fix their output DC level. The additional CMFB circuitry serves the function of controlling the output common mode DC level. The telescopic amplifier designed here using CMFB circuitry is simulated using 180nm technology in Cadence EDA tool. A low frequency gain of 63.9dB is obtained with a power dissipation of 1.38mW and a Unity Gain Frequency of 114MHz.**

***Keywords*: Common-Mode Feedback (CMFB), g_m/I_D method, fully differential telescopic amplifier**

I. INTRODUCTION

High gain differential amplifier is ubiquitous in today's analog circuits owing to its improved immunity to noise, larger output swings and reduced distortion. In particular, an amplifier with high speed and low power dissipation has high utility in various applications like ADCs and DACs. However, the design of a fully differential telescopic amplifier is complicated as the DC level of its output is not well-defined. One of the reasons for this is the current mismatch [1] between p-channel and n-channel MOSFETs used in the amplifier. To solve this problem and to clearly define the DC common mode level at the output node a CMFB circuit is inevitably needed.

Fig. 1 shows the simplest possible fully differential amplifier with CMFB. In order to clearly define its common-mode level; the required CMFB block is shown which has three inbuilt components, namely, *sense circuit*, *auxiliary amplifier* and *return mechanism* [1]. The sense circuit is implemented using source followers (SFs) and sensing resistors. The sensing is done at the source of the SF by means of resistors of appropriate values [1]. The voltage thus sensed is compared with the desired output DC common mode level by means of an auxillary amplifier. This amplifier regulates the gate voltage of the tail current transistor which comprises the return mechanism. The regulation of the current through the transistor achieves the desired function of controlling the output common mode level. But, simply attaining the desired common-mode level is not enough as the CMFB loop introduces stability issues [5] and hence, the circuit must be compensated accordingly [2,3]. The details of design mechanism follow in the next section.

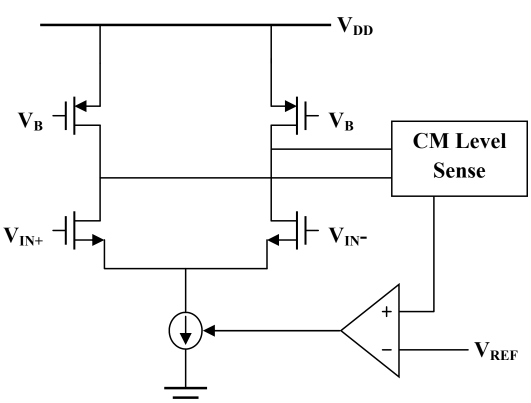


Figure 1. Conceptual topology for Common-Mode Feedback

II. DESIGN METHODOLOGY

A. *Estimating design parameters*

The design is done keeping in mind a set of target specifications. We aimed at a gain of about 60dB (1000V/V), -3dB frequency of 100 kHz, Unity Gain Frequency of 100 MHz and power dissipation of 1mW for the circuit. Now, the low frequency gain (A_v) for the schematic shown in Fig. 2 is $|A_v|$

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_46, © Springer Science+Business Media B.V. 2010

$=g_m*R_{out}$. Also, $\omega_{-3dB} \approx 1/(R_{out}*C_L)$, so we can estimate R_{out} [$1/(\omega_{-3dB}C_L)= 1/(2\pi*100kHz*1.1pF)$] as 1.45MΩ. Hence, g_m can be approximated as [A_v/R_{out} = 1000/1.45MΩ] 0.6897mΩ^{-1}. Now I_D, the bias current of the telescopic amplifier, can be approximated as $P_D/V_{DD}*n$, where P_D (~1mW) is the power dissipation, V_{DD} (3.3V) is the supply voltage and 'n' (=8) is the number of branches in the circuit. The value of I_D comes out to be around 38μA. So, we choose a value of $I_{SS}/2$ as 50μA while currents are mirrored in the remaining branches with appropriate aspect ratios. Having got g_m (~0.6897 mΩ^{-1}) and I_D = $I_{SS}/2$ (~50 μA) and by using the $\mathbf{g_m/I_D}$ technique, V_{OV} for different MOSFETs are found out. Now, by using the I_D/(W/L) plots, W/Ls are determined.

B. *Generating Biases*

In the schematic (Fig. 2), the MOSFETs are numbered left to right and top to bottom (e.g. MOS at top left corner is M_1 while the one at bottom right is M_{31}). In the figure, M_{11} and M_{16} are biased by M_{19} (diode-connected) which is stacked over M_{20} (diode-connected). This puts an upper limit on ICMR of 1.8 V while the lower limit is 1.2 V (V_{T12} + V_{OV13}). Hence, the input bias at M_{12} and M_{17} is chosen midway at 1.5 V. Also, to maximize the output swing, p-channel MOSFETs (M_9, M_{10}, M_{14}, M_{15}) are biased using the low-voltage cascode technique [1].

C. *Sensing the Common Mode Level (V_{CM})*

For sensing V_{CM} we use Source Followers (SFs) followed by resistors. Direct sensing by resistors (R) is not advisable as it affects the gain of the amplifier and necessitates a huge value of R, thereby consuming substantial chip area. However, the use of SF puts a lower limit of 1.3V on OCMR. Moreover, the value of resistors is chosen large enough so that the MOSFETs are not starved during a large signal swing [1]. These resistors can be easily replaced by means of MOSFETs operating in triode (linear) region.

D. *Auxiliary Amplifier*

To compare V_{CM} with V_{REF} (desired output DC level), an auxiliary amplifier is designed with n-MOS loads. The gain of this amplifier need not be very large as a significant gain is provided by the telescopic amplifier for the CMFB loop. Besides, with a high gain auxiliary amplifier, sufficient Phase Margin for the CMFB loop is difficult to achieve. An additional SF stage (using M_{30} and M_{31}) is designed so that ($V_{REF} - V_{GS}$) can be directly compared with ($V_{CM} - V_{GS}$).

E. *Return Mechanism*

For stability and compensation considerations [2], bias current is controlled using two MOSFETs of equal W/Ls (M_{13} with a fixed current and M_{18} with a variable current). The output of the auxiliary amplifier is sensed by M_{18}, which in turn regulates the bias current flowing through the branches of telescopic amplifier. To illustrate, if V_{CM} is greater than V_{REF}, auxiliary amplifier's output increases. This causes tail current through M_{18} to increase and thus, resulting in a drop in the output DC level till it equals V_{REF}.

III. RESULTS

This section outlines the detailed results obtained after performing appropriate simulations for the designed circuit. It is observed that the targeted specifications are met with reasonable accuracy. Some of the specifications are:

ICMR	1.2V – 1.8V
OCMR	1.3V – 2.9V
OUTPUT SWING	$\pm$ 1.6V
POWER DISSIPATION	1.38 mW
SLEW RATE	45.45 V/μsec

TABLE I. DESIGN SPECIFICATIONS

DESIGN SPECIFICATIONS	VALUES OBTAINED AT DIFFERENT PROCESS CORNERS		
	TT (25^0 C, 3.3V)	FF (0^0 C, 3.63V)	SS (100^{0}C, 2.97V)
Amplifier DC (Open Loop) Gain	63.9 dB	53.6 dB	63.6 dB
Phase Margin for amplifier	87.58^0	86.13^0	85.7^0
CMFB Loop Gain	54.97 dB	52.75 dB	45.36 dB
Phase Margin for CMFB Loop	82.05^0	82^0	87.25^0
Unity Gain Frequency	114 MHz	143.5 MHz	86.8 MHz
CMRR	88.4 dB	70.24 dB	72.2 dB
PSRR	103.2dB	87.82 dB	77.98 dB
Settling Time ($\mathbf{t_{1\%}}$)	63 ns	49 ns	79.5 ns
Input Noise (at 10Hz)	3.2 μV/$\sqrt{Hz}$	2.4 μV/$\sqrt{Hz}$	4 μV/$\sqrt{Hz}$

TABLE II. ASPECT RATIOS OF THE TRANSISTORS

MOS	W/L (μm/μm)	MOS	W/L (μm/μm)
M_1, M_5	6.1/0.36	M_{10}, M_{15}	175/0.36
M_2	3.1/0.36	M_{11}, M_{13}, M_{16}, M_{18}	11.25/0.36
M_3	25/0.36	M_{21}, M_{22}, M_{23}, M_{24}	11.25/0.36
M_4, M_7, M_8, M_{19}, M_{20}	1.125/0.36	M_{27}, M_{29}, M_{30}, M_{31}	11.25/0.36
M_6	17.5/0.36	M_{12}, M_{17}	11.25/0.36
M_9, M_{14}, M_{26}, M_{28}	61/0/.36	M_{25}	122/0.36

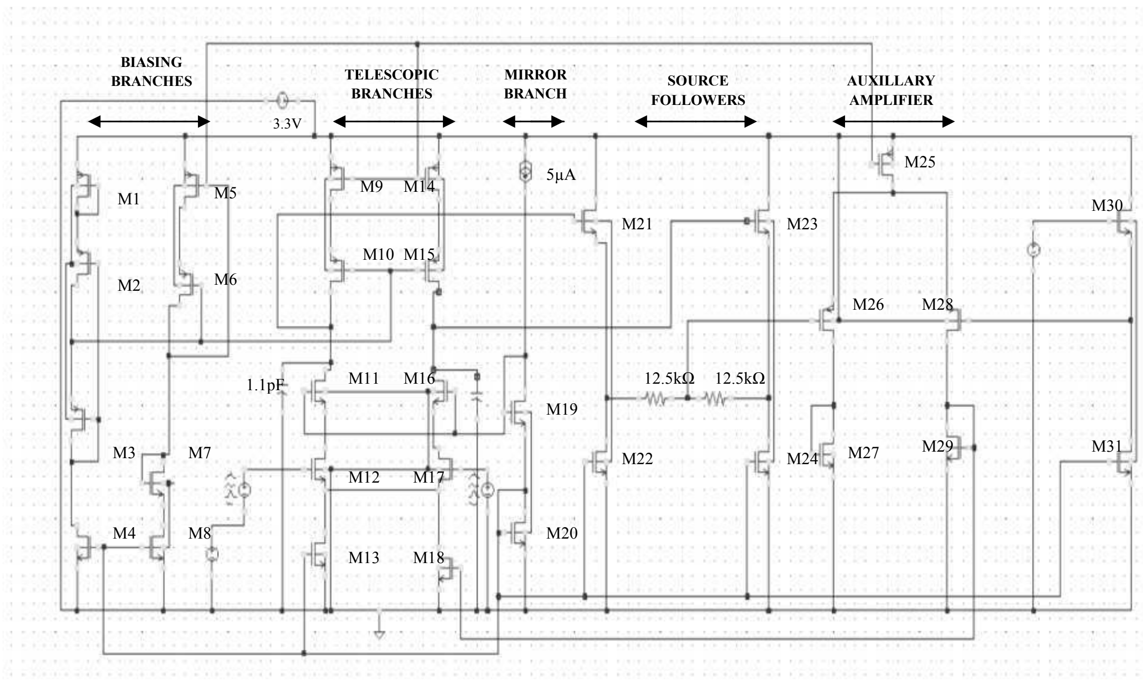


Figure 2. Schematic of Telescopic Cascode Amplifier with CMFB

The remaining specifications are tabulated in Table I while Table II gives the aspect ratios of the MOSFETs. The simulation results obtained for AC analysis (Fig. 3 & Fig.4) and transient analysis (Fig.5 & Fig. 6) are given below.

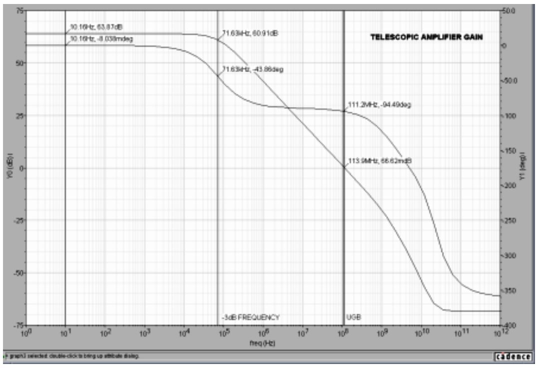


Figure 3. AC gain and phase for the telescopic amplifier

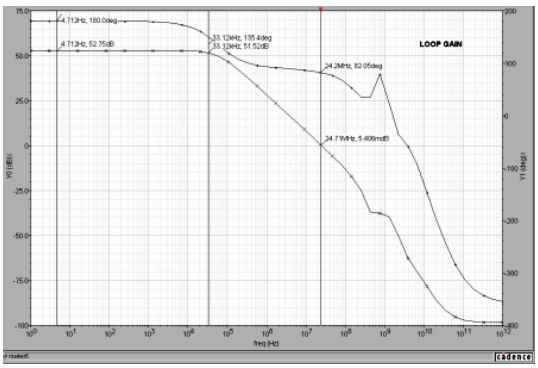


Figure 4. AC gain and phase for the CMFB loop

IV. CONCLUSIONS

A high gain, low power, fully differential telescopic cascode amplifier is implemented using CMFB circuit. Stability of both amplifier as well as CMFB loop [5] is ensured by obtaining sufficient Phase Margins. The potential application of the circuit is in precision ADCs/DACs, Buffers, Comparators, Sample and Hold Amplifiers etc. due to its high gain and bandwidth. Also, the design methodology and CMFB circuit used here can be replicated in the implementation of folded cascode and gain boosted telescopic amplifiers depending upon the specific voltage swing and gain requirements. The only apparent limitation of the circuit turns out to be its marginally low output voltage swing. Thus, the utility of the design can be further augmented by improving the differential output swing and reducing settling times by leveraging the substantial Phase Margins present at our disposal.

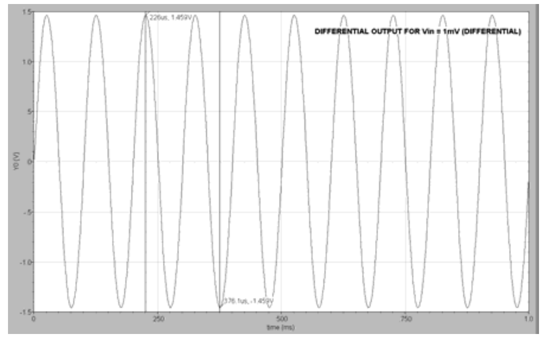


Figure 5. Differential Output for 1mV differential sine input

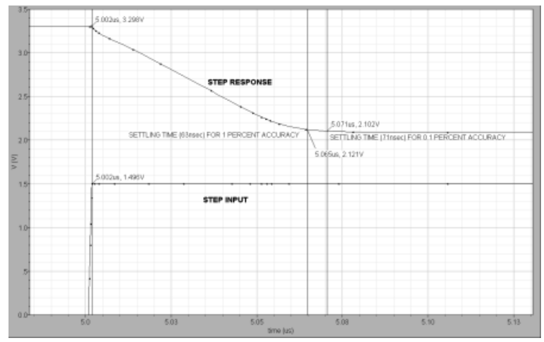


Figure 6. Step response for a step input of 1.5V (settling time 63ns for 1% error)

V. REFERENCES

[1] B. Razavi, Design of Analog CMOS Integrated Circuits, Tata McGraw-Hill Edition:2002.

[2] Paul R. Gray, Paul J. Hurst, Stephen H. Lewis, Robert G. Meyer, Analysis and Design of Analog Integrated Circuits, 4th ed., John Wiley and Sons Inc.:2001.

[3] David Johns, Ken Martin, Analog Integrated Circuit Design, John Wiley and Sons Inc.: Copyright 1997 .

[4] P. M. VanPeteghem, J. F. Duque-Carrillo, "A General Description of Common-Mode Feedback in Fully-Differential Amplifiers," Circuits and Systems, 1990., IEEE International Symposium on, vol.. 4. pp. 320–312, 1-3 May 1990

[5] Tauro, A. Marzocca, C. Francesco, C.T. Di Giandomenico, "Common mode stability in fully differential voltage feedback CMOS amplifiers," A. Dipt. di Elettrotecnica ed Elettronica, Politecnico di Bari, Italy; This paper appears in: Electronics, Circuits and Systems, 2003. ICECS 2003. Proceedings of the 2003 10th IEEE International Conference on Publication Date: 14-17 Dec. 2003 Volume: 1, On page(s): 288- 291 Vol.1

Development of Children's Communicative Competence using Case Method of Technology

Giedrė Strakšienė, Dalia Baziukaitė

Abstract— **The paper addresses the issue of unusual ICT tool application in the development of communicative competences during drama classes. It also presents the practical implementation of a case study method with the application of ICT tools. The case method was used for an interesting combination of acquiring communicative competences including drama skills by means of applying modern computer software. The target group was the 3rd form pupils of Klaipėda primary school. The activities planned to cover skills of language and literature, drama and ICT.**

The significance of the presented case is based on the following: pupils' communicative skills can be improved; learning is based on the integration of inherently different subjects, such as drama, language and ICT; skills of the three subjects are educated in a creative way.

I. Introduction

The case method is a method of learning that is based on active participation, cooperation, and democratic discussion of a practical situation presented in the form of a fiction story or an actual case. A particular case study of technology application was one of the activities of CaMOT (Case Method of Technology for Practical Use of Training Teachers) project[1], supported by the European Commissión Socrates Programme, with Klaipeda University as one of the partners.

The problem is that traditional methods do not provide sufficient opportunities to develop the communicative competence of speaking. Learners train their communicative competence skills, but not to a sufficient degree, as they begin competing and do not try to reach individual and common goals. Communicative competence, and particularly speaking skills, could be best developed through the case study method at primary school.

The main question: does the use of the case method (by integrating ICT, language and drama) help to develop communicative competence? Following the question, the main goal will be to prepare a program of the experiment and verify the program.

Manuscript received October 26, 2008. This work was supported by the European Comisión Socrates Programme, Comenius Action 2.1. Training of School Education Staff 226712-CP-1-2005-ES-COMENIUSC21.

G. Strakšienė is with Department of Pedagogy, Klaipėda University, Herkaus Manto str. 84, 92294, Lithuania (email: giedre.straksiene@ik.ku.lt).

D. Baziukaitė is with Department of Computer Science, Klaipėda University, Herkaus Manto str. 84, 92294, Lithuania (phone: 370-46-398879, fax: 370-46-398822, email: dalia.baziukaite@ik.ku.lt).

[1] http://www.camot.net

With the case study, we intended to show that learning a language is not only a process of learning grammar, but also a social one. Arts and creative activities can be viewed as study subjects in social-educational sciences. New perspectives of linkage between drama, language learning skills, and information communication technologies (ICT) skills were an area of our interest.

The activity took place in a primary school. Two teachers (the form teacher and the drama teacher) were guiding case activities. The pupils involved came from Forms 3 to 4 of the primary school. Primary school in Lithuania consists of 4 forms (from the 1st to the 4th). In the school which was selected for the case study activity pupils are taught various subjects, starting from the main subjects, such as mathematics, language, and literature, and ending with art subjects.

The activities planned in the case covered skills from language and literature, drama and ICT. Language and literature skills are mainly listening, writing, and creating dialogs. Drama skills are recognition of emotions and mimic and improvement of speaking quality. ICT skills are typing a text, managing a standard voice recorder, and manipulating other standard computer software.

The case was divided into five steps. In the first step of the case, children read materials and watched photo-mimic presentation. In the second step, pupils had to solve several types of tasks, e.g., to find a matching description of an emotion, or to find a suitable description for a given emotion or mimic. In the third step, pupils were presented with a recorded video situation in which two peoples were communicating.

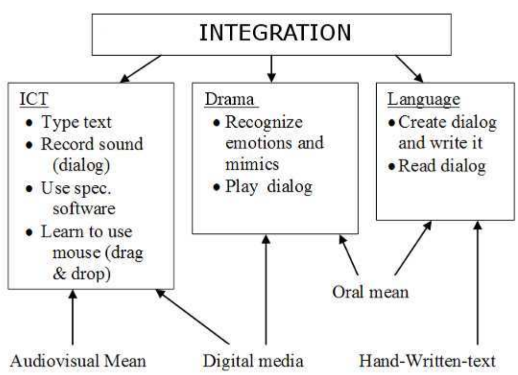


Fig. 1. Trained communicative competences and their relation

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_47, © Springer Science+Business Media B.V. 2010

The film was shown without sound. Pupils then could write dialogs in the next step. The fourth step included dialog creation and their recording with the help of a standard voice recording software and typed texts. The fifth step was devoted to the presentation of created dialogs. Teacher collected all the created dialogs and played them one after another. At the end, the film was played with the original sound, and the results were compared.

The communicative competences depicted in Fig. 1 were trained by the case activities. The significance of the presented case was based on the following:

- pupils' communicative skills were improved;
- learning was based on the integration of inherently different subjects, such as drama, language and ICT;
- skills of the three subjects were learned in a creative way.

The next section gives a short description on the case study method. Section III explains a notion of communicative competences and the role of basic emotions in developing described competences. Section IV presents a tool created specially for practical implementation of the case activities in school. Section V analyses pupil achievements during an experimental lesson at school. Section VI gives suggestions on how the created tool and case study method can be used in training teachers. Finally, we conclude with a summary of results and our vision for future work.

II. Case Study Method

The students can learn more effectively, if they are actively involved in the learning process. In science theory the case study approach is one of the ways to implement active learning strategies in schools. According to C. Davis and E. Wilcock [17] there are a number of definitions for the term "case study". Fry et al. described case study as complex example which gives an insight to the context of a problem as well as it illustrates the main issue. Fry defines case study as student activity based on topics that demonstrate theoretical concepts in an applied setting. This definition of case study covers a variety of different teaching structures and methods, ranging from short individual case studies to longer group-based activities. Case study is considered to be useful pedagogical tool. Also they noticed, the benefits of using case studies as an interactive learning strategy, shifting the emphasis from teacher-centered to more student-centered activities that Grant has outlined. The important fact is that case studies are linked with increasing student motivation and interest in a subject [13], [17].

Agreeing with authors C. Davis and E. Wilcock, we think that case studies are useful, because they:

- allow to demonstrate theoretical concepts in application, thus bridging the gap between theory and practice;
- encourage active learning;
- provide an opportunity to develop key skills, such as communication, working in groups, and problem solving [17].

The case method is based on a learning principle that states: people learn best when they teach themselves through their own effort. One can gain greater understanding and improve skills in judgment when one works through a problem, but not just passively listens to a lecture. Similarly, there will be greater learning results if someone "uses" a theory in practice.

Cases can help us in learning to:

- apply theories to real situations;
- solve real problems.

Like real situations, cases center on an array of partially-ordered, ambiguous, seemingly contradictory and reasonably unstructured facts, opinions, inferences and bits of information, data, and incidents out of which someone must provide order by selectively choosing which bits to use and which to ignore. In real life, others will not do this for us. Similarly, in real life situations, it is unlikely that any two peoples would assemble the data or make inferences identically. Someone will have to work within the limitations inherent in evidence and arrive at internally consistent interpretations. Experiencing the process of learning this way may be frustrating and confusing, but it is also practical and realistic.

III. Basic Emotions within Communicative Competences

Communicative competence:

- is the ability to get what you are seeking from others in a manner that maintains the relationship on terms that are acceptable to all parties [1];
- does not mean behaving the same way in all settings and with all people [4];
- competence varies from one situation to another [16];
- is an aggregate of individual abilities, knowledge and experience that conditions self-confidence and correct orientation when communicating with each other [10];
- communication competence is, to a high degree, a set of skills that can be learnt [1].

Authors, who address communication processes, essentially agree that communication is continuous, dynamic, and shifting process. Researchers accept that communication requires specific personal capabilities. A capability which according a qualification, knowledge, and abilities allows particular person to perform communicative activity is defined as communicative competence.

M. Canale distinguishes features to characterize communication and states that communication is a part of socio-cultural context, a form of social interaction, which is closely related with language. Expression forms of language in most cases are unpredictable and are creativity based.

Real Decreto 1006/1991 of 14th of June establishes the teaching requirements for primary education nation-wide. In it, communicative competence comprises five sub-competences: grammar competence, sociolinguistic competence, socio-cultural competence, discourse competence, strategic competence. For D. Hymes [7] the communicative competence has four aspects: systematic potential, appropriacy, occurrence, feasibility. These categories have been adapted for teaching purposes.

The majority of students understand emotions as feelings that come up in different situations. Another group of students explain them as being "what the body tells us". To describe feelings is a rather difficult task to students. Several of them noted, that generally feelings are apparently observable in physical changes; one of them involve a person's appearance (for instance, blushing, sweating) and other changes involve behavior (for instance, facial expression gestures, posture, and so on). Body language is an important part of communication using movements or gestures and is used consciously and subconsciously to convey messages. Teachers-pupils and pupils-pupils constantly send and receive nonverbal messages many times a day. Pupils need to be aware of how to use gesture, posture, facial expression and tone of voice effectively to establish a good relationship.

Social scientists generally agree that a phenomenon known as "feeling" consists of several components: physiological factors, nonverbal reactions, cognitive interpretations and verbal expression [1].

In fact, it is impossible to talk about communication without acknowledging the importance of emotions. The role of emotions in human affairs is apparent to social scientists and laymen alike. According to psychologists R. J. Sternberg [15] and D. Goleman [6] there is such a phenomenon as "emotional intelligence".

D. Goleman [6] makes the claim that intellectual ability is more than a way to measure one's talents and that success in the world greatly depends on the ability to understand and manage one's own emotions, as well as to empathize. D. Goleman [6] maintained that "Emotional Intelligence refers to the capacity for recognizing our own feelings and those of others, for motivating ourselves, and for managing emotions well in ourselves and in our relationships", e.g., every thought, gesture, muscle, tension, everything is significant and meaningful, and related to the present. According to R. Sternberg and D. Goleman, it is possible to know and understand oneself on all these levels, and the more one knows the more he/she is free to determine their own life.

Generally speaking, the types of emotion are: primary and mixed emotions, mild and intensive emotions. Famous psychologist R. Pluchik [9] has identified eight primary emotions (joy, acceptance, fear, surprise, sadness, disgust, anger, anticipation) and those primary emotions can combine to a mixed form (optimism, love, submission, awe, disappointment, remorse, contempt, aggressiveness). In real situations, we would probably feel mixed emotions.

Some researchers believe there are several "basic" or "primary" emotions. However, there isn't much agreement among scholars about that those emotions are, or about what makes them "basic". Moreover, emotions that are primary in one culture may not be primary in others, and some emotions have not direct equivalent in other cultures. Despite this, most of them accept that anger, joy, fear and sadness are common typical emotions [1]. Emotions can be sorted according to different degrees of intensity state or arousal levels. They play an important role in all types of human everyday relationships.

During the experiment, pupils had to recognize emotions from photos taken from their own school life. This is a way to provide information to students on how emotions are expressed in speech. The task was to create an audio clip in accordance with emotional situations presented in the photos (see next section for details).

IV. A Tool to Help Learning Emotions

We created a simple html based tool, which we have used during the experiment at the school. It consisted of a collection of materials (photos, videos, text, quizzes) that pupils could learn from, or at least looked through, trying to understand different types of emotions, mimics and their importance in human communication. A tool, as it was used during the experimental lesson, is available for viewing online[2] in Lithuanian language. It contains all five stages that train different types of skills (Fig. 2) included in the presented case study.

The stages of our qualifying case were as follows:

1. *Discover.* Texts and simple graphics or even video material are presented.
2. *Apply.* Solve several types of tasks.
3. *Observe.* Pupils are presented with recorded video situations in which two people communicate.
4. *Create.* Write dialog, type it, record it.
5. *Present.* Present created dialogs.

In the first stage of the case, two types of emotion description were presented to pupils. First, in the written form, they were presented with the structure of three rows for each emotion - the title, a photo expressing the emotion, and a short story describing possible feelings and thoughts captured in the photo (Fig. 4). Then, they watched slides with recorded explanation of the emotions expressed in the previously presented photos. The second stage was performed with quizzes that were created with the Hot Potatoes tool[3]. During this step pupils had an opportunity to test their previous experience and possibly apply the newly acquired knowledge from the first stage of the case study.

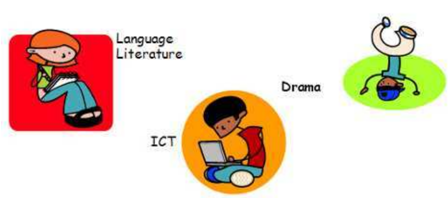


Fig. 2. Trained skills

[2] http://ik.ku.lt/~dalia/camot
[3] http://hotpot.uvic.ca

Fig. 3. Using specially prepared materials

Fig. 4. Working in pairs

Fig. 5. Creation of dialog

Fig. 6. Listening self recorded track

The third stage was supported with example dialogs. The dialogs were created in accordance with the situation expressed in the photo, coming together with a particular dialog. There were three dialogs. The activity of the third stage helped pupils to prepare for the next stage, in which they created their own interpretations.

During the fourth stage, children created a text in accordance with a selected picture, wrote it down on the paper by hand, and afterwards typed the text using the text processor (Fig. 3, 5). For that activity, the tool included photos and video with different situations. After typing the dialogs in a text processor, pupil used standard audio recorders that came together with an operating system, and recorded their dialogs.

The fifth stage was devoted to the presentation of the dialogs (Fig. 6) created in accordance with the given situations expressed in the photos or video materials.

All the results of the practical implementation of the case study were collected and presented in the CaMOT project website[4] under the case entitled "Learning to recognize emotions. A case in primary school".

V. Application of Case Method

Social skill (abilities) have significant role in relations with other peoples as with members of a society. That is why it is important to develop communicative competence starting from the youngest age at primary school. Teachers should not neglect the social aspect in language learning, too. Teachers are trying *to find new ideas, materials and methods how to develop communicative skills*, however, they most often tend to use traditional methods [16]. Language skills are necessary for displaying language proficiency and social abilities. Language is action with which we start and maintain relationships.

The educational reform ideas in Lithuania and experience of educators from Western countries stimulate the active use of various training methods in Lithuania. Nowadays teachers have a good opportunity to teach, train learners in nontraditional ways.

The students of the third form are at the verge of adolescence, and their interests are diverse, as they start seeing the world in different ways. Various aspects become important to them, and the teacher's role is to keep them interested and help them to develop in a variety of ways. Social skills are important and are not to be neglected.

Modern program on primary education is oriented to the development of pupils' competences. The content of its development does not eliminate subjects, but helps teacher to better understand that the goal of the educational process is a child with developed personal competences, but not a subject that should be taught by a teacher and learned by a pupil [3]. Talking about communicative competence and their development, the emphasis is placed on communication and creativity. That is due to the impact of the human communicative action during life: the readiness to communicate, which means an ability to express one's own thoughts, feelings, opinion, and emotions in different situations, to transmit information in a precise way, to have an ability to create various types of texts, to be polite and follow cultural standards.

Communicative competence is a complex psychological, pedagogical, social and physiological process that allows a person to interact, to send and receive information (in all oral, written, gesture and mime forms) and to become a full member of the knowledge society. In order to develop communicative competence in a primary school, new methods and forms of teaching should always be under consideration. For certain goals, tasks in game format are usually used, known as didactic games, or it can be practical tasks of writing and reading during language lessons. Development of technology and new computer-based game possibilities make an impact on art specialists as well, making them think about possible application of ICT in their subjects. Until now, ICT is rarely used as a tool for drama tasks in Lithuania.

In every case of a dialog, a subtext (what it is hidden under the pronounced text) comes out. Participants of a dialog can hear the intonation and feel the emotive state of a person, can decide on their temperament and other things that can only be seen in communication. Next to the verbal language, a language of a different dimension exists, and it is called non-verbal language. Often it provides more information than the first one and influences the verbal language. Learning to talk by using the opportunities of non-verbal language is important from primary school. The combination of verbal and non-verbal expression makes a message clear, suggestive, meaningful, original, etc. An important moment is how we are able to manage the content of our language, formulate ideas and express them in a clear way, how we manage our mimic and gesture and how we make our message more suggestive. As it turned out, pupils in the third form could recognize basic emotions rather well. Role-playing games were very useful and contributed to the pupils' getting engrossed in the text and their understanding of its content and emotive background. An interesting task for pupils is the creation of a sound track.

[4] http://www.camot.net

VI. Pupil Achievements

As explained in the introductory section of the paper, pupils from one of Klaipėda secondary schools took part in the experiment. The experiment was carried out in May, 2007. 26 students from the 3rd form of primary school took part in the planned activities.

After the pupils introduced their selves with theoretical background of emotions, several quizzes were put in front of them. Quizzes were prepared in a simple way. A photo expressing some emotion was displayed on a screen (3 photos / screen). Pupils had to choose one of three possible answers to describe the emotion given in the picture. Students had to pass through four such screens with photos and possible answers. The computer program did the calculation of scores for one screen result. Quizzes were prepared with Hot Potatoes[3] software. Students had the possibility to improve their answer in case the first attempt was with mistakes, but the final score was computed according primary result the student got in the first attempt.

Table 1 summarizes the results of quiz answers. The columns of the table represent a number of pupils who answered the quiz questions with 100, 66, 33 and 0 percentage of quality (33% means that only one emotion given on the screen was recognized, 66% two emotions were recognized, 100% means that the task was completed without mistakes). In section IV, every stage of the case study was explained. Some conclusion can be made after summarizing the results of the answers to the quiz questions, in which they had to recognize emotions (from the stance, mimics, gestures, signs) that were shown in the photos. Afterwards, the pupils tried to identify the emotions in the photos. The summary of the results of Table 1 is presented in Fig. 7; in accordance with it:

- Always identifies correctly (100%) - 32%
- Often identifies correctly (66%) - 35%
- Sometimes identifies correctly (33%) - 31%
- Never identifies correctly (0%) - 2%

Coming from the results of this concrete task, we conclude that the most part of pupils can recognize emotion quite good (Fig. 7.) and interactive ICT tools made the process acceptable and playful.

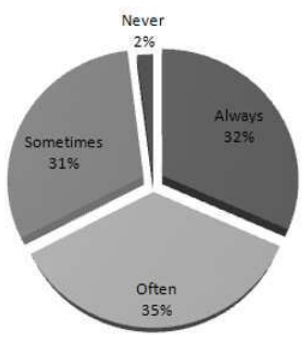


Fig. 7. Answer distribution: 67%of target group was able to identify emotion right

TABLE I

DISTRIBUTION OF PUPIL ANSWERS TO FOUR GIVEN QUIZZES

Number of answers:	100%	66%	33%	0%
1 task	20	2	4	0
2 task	7	9	8	2
3 task	4	15	7	0
4 task	2	11	13	0

Fig. 8 - 13 show some dialogs in their written and typed form. As can be seen in the pictures representing handwritten and typed examples of dialogs, grammar mistakes made in the written forms of a dialog usually remain in the typed version of it. *An interesting finding was that pupils made new ones while typing dialogs into a text processor*. That might be explained as a lack of concentration while doing such a type of tasks at this age. Also, an impact of modern mobile technologies should be under consideration, because it is not a secret that people prefer to write SMSs without using special characters of Lithuanian language.

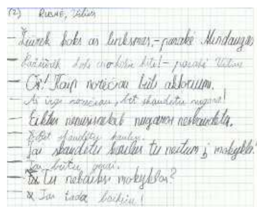
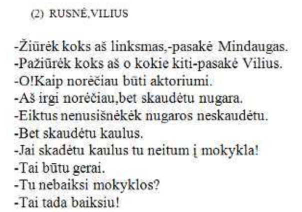


Fig. 8.



(2) RUSNĖ,VILIUS

-Žiūrėk koks aš linksmas,-pasakė Mindaugas.
-Pažiūrėk koks aš o kokie kiti-pasakė Vilius.
-O!Kaip norėčiau būti aktoriumi.
-Aš irgi norėčiau,bet skaudėtu nugara.
-Eiktus nenusišnėkėk nugaros neskaudėtu.
-Bet skaudėtu kaulus.
-Jai skadėtu kaulus tu neitum į mokykla!
-Tai būtu gerai.
-Tu nebaiksi mokyklos?
-Tai tada baiksiu!

Fig. 9.

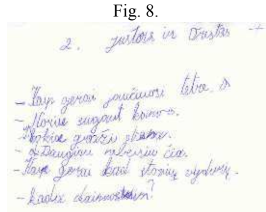
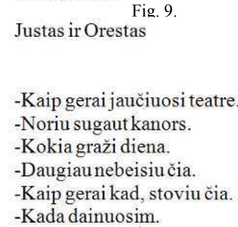


Fig. 10.

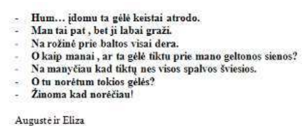


Justas ir Orestas

-Kaip gerai jaučiuosi teatre.
-Noriu sugaut kanors.
-Kokia graži diena.
-Daugiau nebeisiu čia.
-Kaip gerai kad, stoviu čia.
-Kada dainuosim.

Fig. 11.

Fig. 12.



- Hum… įdomu ta gėlė keistai atrodo.
- Man tai pat , bet ji labai graži.
- Na rožinė prie baltos visai dera.
- O kaip manai , ar ta gėlė tiktu prie mano geltonos sienos?
- Na manyčiau kad tiktų nes visos spalvos šviesios.
- O tu norėtum tokios gėlės?
- Žinoma kad norėčiau!

Augustė ir Eliza

Fig. 13.

VII. Final Remarks

The case demonstrated a way how language (written and spoken), drama (moving, mimics, and emotions), and ICT (typing text, managing voice recorder, manipulating standard computer software) can be used in an integrated way for teaching and acquiring communicative competences. Such a combination of tools allows reaching an infinite capacity to express and understand the meaning of different types of communicative phenomena.

The main focus of the lesson was on the pupils' ability to create dialogs on the basis of different emotions they identified, given the fact that all the activities were carried out by using modern ICT tools and software.

Traditional approaches treat a language as a free standing package of knowledge, used to analyze and observe it. Many of us have learnt language that way. Learning a language is a very abstract process, and it is known that the traditional approach to learning it does not appeal to everyone. The best results of learning a language are seen when learning becomes a part of our life experience.

One of the most common uses of language for the expressive or emotional purpose is a means of getting rid of our nervous energy, when we are under stress, or are happy and joyful, angry, afraid, etc. Traditional methods of teaching do not give enough opportunities to develop oral communicative competence.

Combined tools incorporating advances of ICT bring new possibilities to methods used for training communicative competences and find their place even in teaching arts and are successfully applied as it was shown in this particular case.

The particular case study can be extended to a broader activity by attracting subject teachers and developing special software tool and teaching methodology.

VIII. Acknowledgments

Most of this work was done as activities of CaMOT (Case Method of Technology for Practical use of Training Teachers) project[5], supported by the European Comisión Socrates Programme, Comenius Action 2.1. Training of School Education Staff 226712-CP-1-2005-ES-COMENIUSC21 under coordination of the University of Cadiz, Spain.

We thank a community of Klaipėda Vydūnas Secondary School for the ability to arrange the case study of technology and for their kind participation in the activities.

References

[1] Adler, R. B., Proctor R. F. II, (2007). Looking out/looking in, Twelfth Edition. Thomson Wadsworth.
[2] Bennett, M. (1997). How not to be a Fluent Fool: Understanding the Cultural Dimensions of Language. In A. E. Fantini (ed.), New Ways in Teaching Culture. Alexandria: TESOL.
[3] (2003). Bendrosios programos ir išsilavinimo standartai. Priešmokyklinis, pradinis ir pagrindinis ugdymas, Vilnius.
[4] Canale, M. and M. Swain (1980). Theoretical Bases of Communicative Approaches to Second LanguageTeaching and Testing. In Applied Linguistics 1, 1-47.
[5] Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards and R. Schmidt (Eds.), Language and Communication. (pp. 2-27). London: Longman.
[6] Goleman, D. (1998). Emotional Intelligence. Working with Emotional Intelligence, London, Bloomsbury Publishing.
[7] Hymes D. (1972). Models of the Interaction of Language and Social Life'. In: J. Gumperz and D. Hymes, Directions in Sociolinguistics: the Ethnography of Communication. New York: Holt, Rinehart and Winston.
[8] Penman, C. (2005). Towards a holistic evaluation of communicative ability in tandem interactions. In Penman, C. (ed.) Holistic approaches to language learning. Frankfurt: Peter Lang, 187 - 199.
[9] Plutchik, R. (1980). A General Psychoevolutionary theory of emotion. In Plutchik& H. Kellerman (Eds.), Emotion: Theory, research, and experience: Vol.1. Theories of emotion. New York: Academic.
[10] (1993). Psichologijos žodynas / red. R. Augis ir kt. - Vilnius: Mokslo ir enciklopediju¸ leidykla.
[11] Sacks, H. (1992). Lectures on Conversation. Oxford: Blackwell Publishers Ltd.
[12] Maines B. (2003) Reading faces and learning about human emotions. Bristol: Lucky Duck.
[13] Mustoe L. R., Croft A., C., Motivating Engineering Students by Using, Modern Case Studies, International Journal of Engineering Education,, Vol. 15, No 6, 1999 TEMPUS Publications, pp 469-476.
[14] Sivan A., Wong Leung R., Woon C. and Kember D. (2000). An Implementation of Active Learning and its Effect on the Quality of Student Learning Innovations in Education and Training International. Vol. 37 No 4.
[15] Sternberg, R. J. (1985). Implicit Theories of Intelligence, Creativity, and Wisdom. Journal of Personality and Social PsychologyVol. 49, n 3.
[16] Šernas V. (2006). Komunikacinės kompetencijos, jų¸ ugdymo(-si) kai kurios strateginės ir taktinės problemos. Santalka. Filologija. Edukologija, T.14 , Nr.4.
[17] Davis C., Wilcock E., Teaching Materials Using Case Studies http://www.materials.ac.uk/guides/casestudies.asp.

[5] http://www.camot.net

Making the University Admission Decision Making Process Lean

Kai Jin, and Manisha Kachroo
Texas A&M University Kingsville
700 University Blvd.
Kingsville, TX 78363
Kai.Jin@tamuk.edu

***Abstract*- To apply Lean strategy/principles anywhere means doing more and more with less and less – less time, space, machinery, materials and human efforts. Although Lean principles are rooted in manufacturing, they apply universally. The challenge for each firm is to translate, tailor and apply them to its own situation. This paper deals with the deep understanding of the functioning of Office of Admissions at Texas A&M University Kingsville, and the usage of Lean concepts, practices and principles to offer some suggestions and recommendations in order to improve the admission decision making process. An effort has been made to completely understand the work of the office (right from receiving of the application packet from a student to the final admission decision made for that student) and make work flows for each category of students applying for admissions. Some recommendations based on lean principles have given to make the admission decision making process lean and reduce the lead time to evaluate and process the admission applications. The Office of Admissions has been viewed as an organization with its product being the final decisions made for students about their admission to the school.**

I. INTRODUCTION

Many manufacturing and service businesses are using lean management principles and practices as a means to improve business processes, which in turn improves productivity and competitiveness and deliver greater value to end use customers. The office of admissions is one of the vital offices of Texas A&M University, Kingsville. It comprises of eleven staff members. The flow of work starts right from receiving the application packet from a prospective student to the final admission decision made for that student. In this paper the complicated functioning of the admissions office has been highlighted and its simple and smooth work flow is desired. The usage of Lean and its concepts, practices and principles has been done while working on the work flows and while offering suggestions and recommendations. Here the admissions office is the organization in which the Lean can be implemented so that the elimination of waste processing time is done and the customer i.e. the student applying for admissions are benefited.

The underlying aims of the paper are:

1) Understanding the complicated and diverse work flow of the Office of Admissions
2) Making the proper and up to date work flows in each category (domestic, international, graduate, undergraduate, freshman, transfer).
3) Easing out and correcting the work flows.
4) Improving consistency in the working of the office.
5) Eliminating waste in terms of time, efforts, manpower, etc and improving quality and relevance of work done by each employee of the office.
6) Delivering great value as perceived by students / applicants.

II. LITERATURE REVIEW

This paper was highly influenced by the research paper [1] which deals with the application of lean principles and practices to the design and delivery of a graduate business course on leadership taken by part time working professional students in a classroom setting. The principle objectives were to improve consistency between what was taught in the course and how the course was taught, eliminate waste, improve course materials and deliver greater value to the students. The results indicated higher level of student satisfaction, in part through clearer expectations, less ambiguity regarding lectures and assignments, standard formats for assignment, smoothing individual and team assignments over the semester and better management of students' time both in and outside class. The prime principle used was continuous improvement. Applying Lean approach required professors to challenge their views regarding what they teach and how they teach it. Usually they teach in the same ways they were taught and so remain bound to convention due to a lack of critical thinking and despite the existence of compelling reasons for change. Likewise, in our project, the office of admissions was doing their work in the old conventional way and had a single work flow, not much descriptive. Upon continuously improving the work flows, we helped them revise admission workflows five times to reach to their present forms and are still open for any further improvements in future.

The paper [1] also deals with the syllabus which is highly detailed and has several pages. In an attempt to completely detail all aspects of the course, the result can be confusion among students. More pages create more opportunities for differences in interpretation of requirements which can lead to disputes between students and professors or administrators. Also unnecessary variation will require the professor to spend a lot of time with students clarifying matters one-on-one, and may inadvertently give inconsistent or contradictory directions to students. It may make grading more difficult or less consistent. Due to lengthy, complicated and unclear

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_48,

syllabi, the course may lack focus and relevance despite good intentions. According to Lean thinking, simplifying the learning contract and making it less ambiguous, including grading criteria is welcomed as a challenging improvement opportunity. This point of the paper parallels to our project when the factor, namely, design of the admission catalogue is taken into consideration. The catalogue should contain the entire necessary required criterion for applying in the school so that neither the time of the employees of the office nor that of the applicants is wasted. The checklist should be descriptive and simple. For e.g. if a catalogue clearly reveals that an international student should have TOEFL score higher than 71 to apply, automatically students having less score will not apply. Contrary to this if no detail is mentioned in catalogue, less scored applicants may also apply. As a result the office will have to spend time in opening those application packets, typing rejection letters and responding back to students about their rejection. By applying Lean, elimination of time and energy will be realized.

Also in paper [2], authors highlight that the Lean thinking can not only change the way of manufacturing but also can be the way to improve any kind of organization. This paper shows how the lean principles can be used to transform the university system to a lean organization. The three main areas in which lean has been tried to implement are course objectives, course nature and course delivery. Assignments, tests and grading systems have been tried to change for better. After applying lean concepts, students' benefits are given first concern, closer relationship between faculty and student is created, the clear, simple and descriptive information is available online and is easily accessible to students, and personalized assignments are created.

The paper [3] highlights the work performed to develop lean six sigma case study for classroom instruction. A full integrated and self-contained study was developed that would enhance student's learning of the lean six sigma methodology. Deming's Plan-Do-Check-Act cycle is utilized to show the importance of continuous improvement.

The research paper [4] reveals the effective educational administration of the curricula in Mexico, reached by inducing Lean concepts in it. Since there are more than two hundred higher education institutions enrolling more than forty thousand students in Puebla Mexico, it becomes difficult to assure the quality of education. What even adds more to this problem is that economical factor is the main criteria for decisions for most institutions. So quality has to pay the price. In this paper, data from different colleges has been collected by the authors and an attempt has been made to modify the curricula, allowing the curricula administrators to understand this modification and influence, assuring a lean and effective education. Due to the opening of more and more colleges and the authorities taking active part in the rat race for gaining profits, the quality of education has been jeopardized. This has led to an imbalance between educational quality and financial health. The core issue of the research is to harmonically balance educational pertinence and financial sustainability. For this the authors suggest reduction of decision times, costs and organizational layers to increase productivity, quality and satisfaction. The emphasis should be on learning rather than in teaching. Right job should be done at the right time. Prosumer concept has been emphasized in the paper. According to this concept the customer participates in his own service or order fulfillment. The students are asked to fill out questionnaire to know their views, suggestions and satisfaction level. Likewise, we also included this point in our research and suggested the office of admissions to take the views of the students who are currently the students of TAMUK. These students can be good if not the best judges of the functioning of this office since they are the ultimate customers.

III. PROJECT BACKGROUND

The Office of Admissions is a critical department of any university because they are the first department that prospective students will work with. The efficiency of their work will dramatically affect the university enrollment and student retention. This paper is based on a project entitled, "Improving the admission decision making process and work flow of Admissions Office using Lean Principles" sponsored by Texas A&M University Kingsville. The response time to the applicants is the most sensitive factors that most customers, i.e. prospective students, are concerned. Applicants want to know the admission decision from the university in a timely manner. However, when we studied the work of the Office of Admissions we found out that the whole process to make the decision is so complicated and it need cooperation from several departments in the university, such as CIS, academic departments, financial aid department, graduate school, and international office. The workload of the admission office is also huge because of the large number of applications every semester. A standard and smooth work flow is necessary to make the process lean.

The usage of Lean and its concepts, practices and principles has been done while working on the work flows and while offering suggestions and recommendations. Lean is being used by many manufacturing and services businesses to improve the business processes, ultimately delivering greater value to the end-use customers. The lean principles can help in streamlining the work flows. Here the admissions office is the organization in which lean can be implemented so that the elimination of waste is done and the customers i.e. the students applying for admissions are benefited.

IV. MAKING THE ADMISSION PROCESS LEAN

A. *Working of the Office of Admissions and Existing Problems*

The office of admissions makes decisions about the acceptance or rejection of an applicant. The applications can be classified into the following categories:

Freshman/Undergraduate	Domestic	International
Graduate	Domestic	International
Transfer	Domestic	International

The Office of Admissions can be considered as a separate entity or an organization. The product here is the decision about the admission status of a student applying for pursuing education. The customers are the students who want to join the school and want to know about their admission decision. So, like any organization, the Lean concepts can be applied to improve the working of the office, in other words, to improve the admission decision making process. The old work flow for each category of applicants has been developed and then modified to make each process lean. Each work flow has been modified five times to its current status. The staff members of the admission's office involved in the modification work. That gave them opportunities to understand what and how their work is done. Sometimes they could point out what and where are the loopholes by themselves. Some suggestions based on lean principles were offered to the office to enhance the pace of their work.

When an application packet reaches the Admissions Office, immediately the office has to decide whether it is complete or not. If it is complete it has to move further and circulated in the office for processing. Temporarily it can be kept in a rack with proper name tags. If the application is incomplete, place it in separate area/rack having red tags. The staff,(student employees) have to make the list of things missing and immediately email the student about it. Target for this should be within 24 hours. Also a basic eligibility criterion for different types of students (domestic, international, transfer, etc) should be handy for all the staff. On the very first hand whatever the information is available in the application packet, the concerned staff should check if the student is eligible for taking admission in TAMUK. This is because it will be an utter wastage of time in devoting time for such applications. So such students should be immediately emailed that their admission is rejected because they do not fit into the basic or primitive criterion. Now the office has three types of applications:

1) Application of students who are eligible as far as their current information is concerned but more information and documents are needed.
2) Application of students who are not at all eligible and it does not matter whether their application is complete or not.
3) Application of students whose documents are complete and they have a fair chance of getting admission.

Now the office has to work upon the third type of applications because for first type, they have to wait until they get complete information.

Also at this point the office has to decide that second type of students fall under which of the following category:

a) Students who do not meet criterion currently but may be in future they can. For example, international students whose TOEFL score are less than 70 are not currently eligible but they may retake the exam and get qualified score in the future and reapply.
b) Students who do not meet some criterion which cannot be worked upon. For example, a graduate applicant whose undergraduate GPA is lea than 2.0.

For the category a, the students must be immediately mailed stating the reason for their current rejection and they should be encouraged to reapply after meeting the requirements. For the category b, the students should be emailed immediately that they are no more eligible for the school and should not apply again.

The above three categories should be neatly separated from each other and placed in separate racks with distinct colored tags, for example, orange for 1, red for 2 and green for 3.

Improving the cooperation between the Office of Admission and other departments is necessary to make the decision making process lean. Right now, the Office of Admission doesn't have the authority to add applicants to the school's electronic system and make certain change in the student database system. They depend on CIS to do that for them. That waste a lot of time on the communication between the two departments. Also some of the admission decision making will finally made by the academic advisor from the academic department especially for graduate applicants. Right now the process is done by the hardcopy delivery and it waste a lot of time on the miscommunication between the Office of Admission and the academic departments. Sometimes application packet even lost on its way.

There are some last minute applications because university doesn't enforce the application deadline from the enrollment concern. However, staff members spend a lot of time to work on the last minute applications when they supposed to work on the applications for the next semester. Among the existing staff members, only a few should be allotted for those three weeks so that the other staff members can handle the regular applications for the next semester. In this way none of the applications will suffer.

The financial aid priority deadline is in the early of April. So the office receives huge applications, emails, and phone calls during that time. To reduce this, earlier deadline for financial aid in February I recommended. This should be mentioned in catalog also so that the students will be prompted to apply earlier and this will reduce the inflow of documents during March and April.

Right now all the application processing staff answer the incoming calls from applicants and that disturber their work and reduce their work efficiency. A separate employee for external calling answer is recommended. No other employee should be directly accessible to the external calls only if there is any special need. This calling answer employee should be trained in a way that s/he should be able to answer all basic queries of the callers so that no other working staff will be disturbed at most of the time.

B. Recommendations based on Lean Principles

Based on the current analysis of the decision making process of Admissions office and using lean principles System Thinking, PDCA cycles, Continuous Flow, Customer Focus, Zero Muda (waste), Red Tagging, 5-S, and Teamwork, some recommendations were offered . They helped in improving the admission decision making process of the office and they are:

1) Modify the catalogue of Graduate and Undergraduate students for both domestic and international students. After reading the catalogue we realized that it does not contain clear requirements for students to fit in this school. Therefore, the applicants after viewing the catalogue are not in a complete position to understand what exactly is expected by them in order to pursue education at Texas A&M University Kingsville. This in turn wastes a lot of processing time. Incomplete

application packets reach the office and involve a lot of time and energy for being processed by the employees. Due to this, the potential students who fit into the entire criterion face a delay in receiving their admission decision. The application deadline should also be enforced in the catalogues.

2) The work in the staff currently is distributed on the basis of alphabets. One staff processes all applicants with last names starting with the letters A through G, one staff processes all applicants with last names starting with the letters H through P., and the other one staff processes all applicants with last names starting with the letters Q through Z. We realized that this is not an effective way of distributing the work and we recommend the work be distributed as follows. Two persons for processing applications for Undergraduate (both domestic and international) and two persons for processing applications for Graduate (both domestic and international). It will include fresh students as well as transfer students. The new work divisions will allow each employee familiar with one or two working processes and speed up their processing time.

3) In the first letter to the applicants, the estimated processing time should be given so that applicants will know when they expected to know the results. Also the applicants should be contacted periodically before the final decision is issued to inform them their application status. This will reduce a lot of query calls and emails.

4) A better technical training can be given to the employees so that they are familiar with the new work flows.

5) Just like manufacturing organizations, this office can also be considered as an organization with employees working for delivering the product (admission decision) to customers/students as fast as possible. When Lean principles are implemented in a manufacturing organization, it is taken into consideration that the other organizations which are linked in making the product should also be convinced in applying Lean principles and they should modify their work flow at the same time. In the same way, other departments/offices upon which the work of Admissions Office depends should also be requested to give consideration to their work flows. This will help in speeding up the process. Upon interviewing with the staff, we came to know that the most time is consumed when the file is referred to Admission Committee. So that department can also be requested to increase the pace of their work.

6) The work of Admissions office also depends on the CIS which has authorization to some information needed by this office, e.g. the online test reports of GRE. This may in turn prove to be a hindrance in the consistency of the work flow of the admissions office and in turn affect the admission decision making process. This issue can be solved by giving the Office of Admission direct access to the student information system or assigning a particular person from CIS to response the requests from the Office of Admission in a timely manner.

7) Conduct students' survey to find the customers' requirements and further improve current work flow.

8) Using e-files instead of paper based hard copies of application packets when they are transferred between department and that will avoid a lot of miscommunication and file lost. Using automatic processing system to do some of the work, such as name matching and periodic email response.

9) The modifications of the work flow should be considered as an unending process. This is the core concept of Lean thinking. So each employee, anytime, can bring in a novel idea or suggestion which can be given a wise thought.

C. *Flow Chart for the Admission Decision Making Process*

By using lean thinking, the work flow charts for all categories of applications have been developed, but only the flow chart for undergraduate freshman application is shown here in Fig. 1. Those flow charts help the employee to be clear with the whole decision making process and follow the same work flow easily so that the response time to the applicants is reduced. They are also very helpful to the new employee training.

V. CONCLUSION AND FUTURE WORK

The aim of this paper is to highlight the problems encountered by the Admissions Office at Texas A&M University Kingsville, and to provide possible solutions and suggestions based on the Lean Production concepts and principles. The old work flows are created for different categories of applications. After the analysis of the current system and decision making process, new work flows are developed to improve the efficiency of the work and reduce the customer response time. Some recommendations to make the process lean are given. Currently a time study is conducting on the samples from last year's applications. Once it is done, the value stream map of the current process will be developed and further improvement of the work flow will be given.

REFERENCES

[1] M. L. Emiliani, "Improving Business School courses by applying Lean Principles and Practices", *Quality Assurance in Education*, Vol 12 – Number 4, 2004, pp. 175-187.

[2] N. Alp, "The Lean Transformation model for the education system", *Proceeding of the 29th Computers and industrial Engineering Conference*, Montreal, 2001, pp. 82-87.

[3] E. Cudney and R. Kestle, "Lean Six Sigma supply chain case study: Breakthrough in trailer utilization", *EdNet Lean Educators Conference*, Boston MA, April 24-25, 2008 .

[4] M.G. Lopez-Molina and B. Reyes-Guerra, "Measuring lean and effective educational administration of curricula in Mexico", *EdNet Lean Educators Conference*, Boston MA, April 24-25, 2008 .

[5] Pascal Dennis, *Lean Production Simplified*, Second Edition, Productivity Press, New York, 2002.

[6] B. Emiliani, *Real Lean: Understanding the Lean Management System*, Volume one, Productivity Press, New York, 2002.

[7] Lean Blog. Available: http://www.leanblog.org

[8] Lean Enterprise Institute. Available: http://www.lean.org

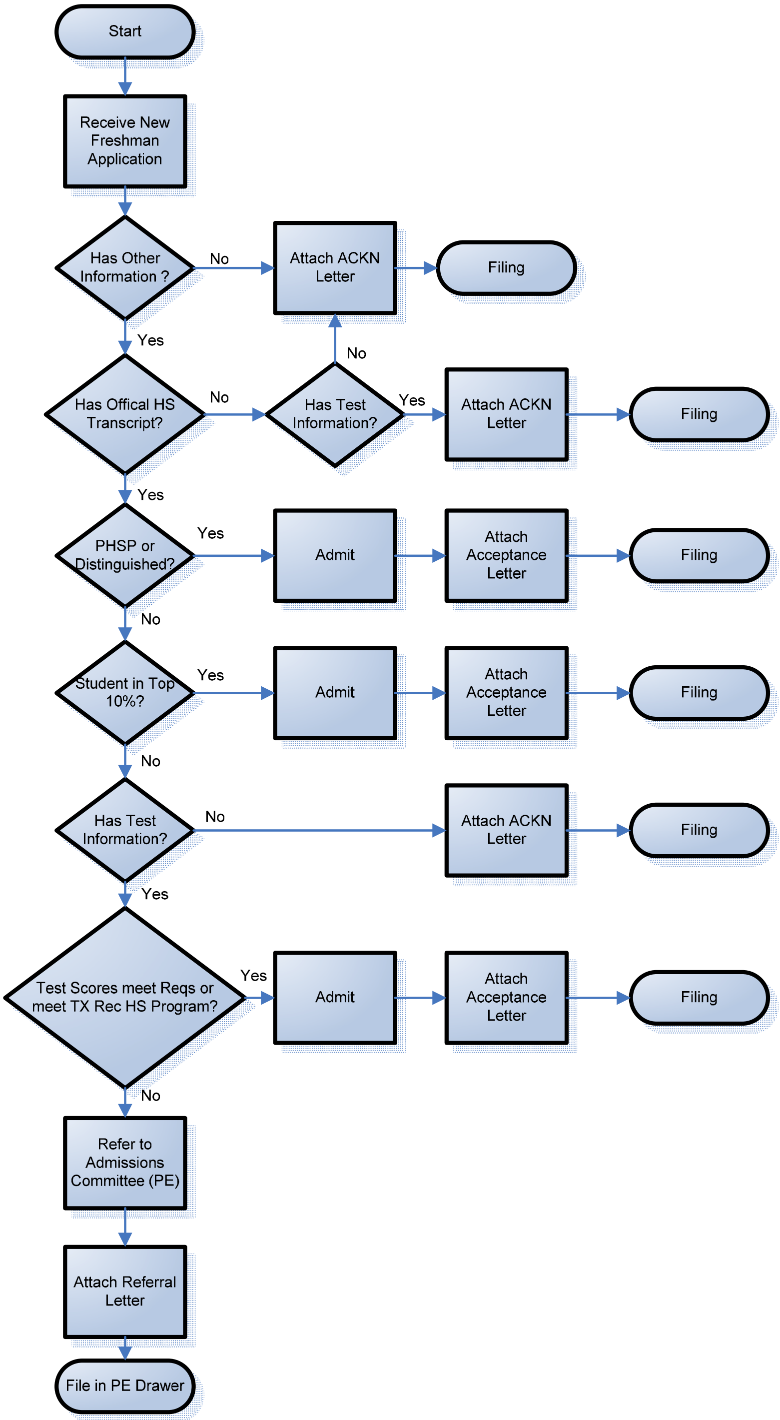


Fig. 1 Flow Chart for the Admission Decision Making Process of Undergraduate Freshman Applications

Control Circuitry for Self-Repairable MEMS Accelerometers

Xingguo Xiong[1], Yu-Liang Wu[2], Wen-Ben Jone[3],
[1] Department of Electrical and Computer Engineering, University of Bridgeport, Bridgeport, CT 06604, USA
[2] Dept. of Computer Science & Engineering, The Chinese University of Hong Kong, Shattin, Hong Kong
[3] Department of ECECS, University of Cincinnati, Cincinnati, OH 45221, USA
Email: [1] xxiong@bridgeport.edu, [2] ylw@cse.cuhk.edu.hk, [3] wjone@ececs.uc.edu

Abstract— **A BISR (Built-in Self-Repairable) MEMS comb accelerometer with modularized design has been previously reported. In this paper, the differential capacitance sensing circuitry for MEMS comb accelerometer is discussed. The BISR control circuitry based on CMOS transmission gates (TGs) is proposed. Each BISR module is connected to the capacitance sensing circuitry through a transmission gate. By turning *on* or *off* a transmission gate, the corresponding module can be either connected to or isolated from the capacitance sensing circuitry. In this way, the faulty module can be easily replaced with a good redundant module for self-repair. The parasitic model for the BISR control circuitry is also analyzed. The analysis results show that the parasitic capacitance will not affect the proper operation of the BISR control circuitry. Furthermore, the signal strength will not be degraded due to the insertion of analog multiplexers. The control circuitry can effectively isolate the faulty module of the BISR MEMS comb accelerometer. Both BISR and non-BISR MEMS accelerometer designs are suggested and their performances are also extracted for comparison.**

Keywords: MEMS (Microelectromechanical System), redundancy repair, parasitic model, capacitance sensing, accelerometer.

I. INTRODUCTION

MEMS comb accelerometers have the advantages of small size, low cost, low energy consumption and better compatibility with VLSI technology. They have been widely used for many applications such as automobile airbag deployment system, aerospace inertial navigation, consumer products [1]. MEMS comb accelerometers generally contains large number of repeated comb finger groups in a very compact manner. For example, an ADXL50 accelerometer contains 42 differential comb finger groups [2], while an ADXL150 accelerometer contains 54 differential comb finger groups [3]. Such a highly dense comb structure with many long and narrow capacitance gaps is extremely vulnerable to various defects such as particle contamination, stiction [4]. Taking the ADXL50 accelerometer as an example, the length of each movable finger is 120μm, while the capacitance gap between each pair of fixed and movable fingers is only 1.3μm. If a conductive particle with diameter larger than 1.3μm falls into any of the 84 capacitance gaps, it will lead to short-circuit of the device capacitance and result in a failure of the entire device. Thus, a large number of finger groups unavoidably lead to the decrease in yield as well as reliability. Furthermore, for some safety-critical applications (e.g. aerospace, biomedical device), the failure of a small MEMS device may lead to catastrophic disaster or the loss of human life. For such applications, extremely high reliability of MEMS devices is required. In order to improve the yield and reliability of MEMS comb accelerometer devices for safety-critical applications, we proposed a built-in self-repairable (BISR) MEMS comb accelerometer device design [5]. As shown in Figure 1 [5], the main device of the comb accelerometer consists of *n* (n=4) identical modules, and *m* (m=2) modules are introduced as the redundancy. If any of the working modules in the main device is found faulty during a built-in self-test (BIST), the control circuit will replace it with a good redundant module. In this way, the faulty device can be self-repaired through redundancy. Due to the redundancy repair, both the yield and reliability of the MEMS comb accelerometer can be improved [5][6].

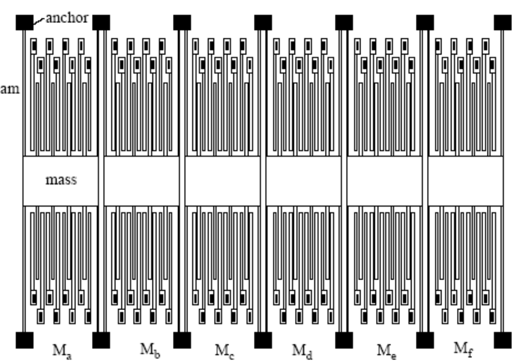


Figure 1. BISR MEMS comb accelerometer design based on redundancy repair

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_49, © Springer Science+Business Media B.V. 2010

The MEMS comb accelerometer relies on a high-resolution differential capacitance sensing circuit to sense the capacitance change due to input acceleration. Furthermore, a BISR control circuitry is needed for the self-repair of the BISR accelerometer. Whenever a faulty module is found in the self-test, the control circuitry should be able to immediately isolate the faulty module and replace it with a good redundant module. In this way, the whole device is still ensured to work properly. In order to control which modules will be connected as main device, an analog switch (transmission gate) is inserted between each individual module and the capacitance sensing circuitry. By turning *on* or *off* these transmission gates, each module can be either connected to or isolated from the main device. Since the differential capacitance change of MEMS comb accelerometer is generally very small (in the range of fF), the parasitic capacitance introduced by the BISR control circuitry cannot be ignored. It is important to find out how the parasitic capacitance will affect the operation of the BISR device, and whether the faulty module can be isolated effectively once it’s identified. In this paper, both the BISR control circuitry and its parasitic model will be analyzed.

II. Circuit Support for BISR MEMS Accelerometer

2.1. Differential Capacitance Sensing Circuit

Various circuit schemes [7]-[11] for signal detection of MEMS differential capacitance sensors have been reported. Among them, capacitance measurement by sensing the current flow through a transducer [11] is very attractive. Due to the high-frequency probe signal utilized, very high sensitivity and speed can be achieved. Further, this circuit scheme is especially suitable for BISR MEMS device, since it ensures better separation of the faulty modules in the BISR device. By turning off the corresponding analog switch, the faulty module will have no contribution to the total current sensed by the OPAMP. Hence, the load effect of the faulty module to the main device can be eliminated. The circuit diagram [11] is shown in Fig. 2.

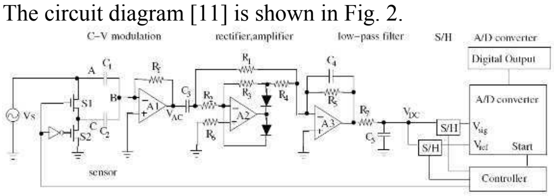


Figure 2. The differential capacitance sensing circuit.

As shown in Figure 2, the sensing circuit [11] consists of five stages: capacitance-to-voltage (C-V) modulation, rectifier and amplifier, low-pass filter, S/H, and A/D converter. A high frequency (1MHz) carrier V_s is used for signal modulation. The differential capacitance of the accelerometer is denoted as C_1 and C_2 separately. The switches S1 and S2 stay in opposite states (on or off) alternately, so that the signal detection circuit senses capacitances (C_1+C_2) and (C_1) in a time-sharing scheme. Assume the OPAMP and switches are ideal, when S1 is on and S2 is off (phase 1), capacitance C1 and C2 are connected in parallel to the voltage source Vs. The output voltage of OPAMP A1 at phase 1 (denoted as V_{AC1}) is:

$$V_{AC1} = -j\omega(C_1 + C_2)R_f V_s$$

This voltage will be converted into a dc signal and applied to the A/D converter as the reference voltage V_{ref}. When S1 is off and S2 is on (phase 2), the capacitance C1 is connected to Vs and C2 is shorted to ground. The output voltage of OPAMP A1 at phase 2 (denoted as V_{AC2}) becomes

$$V_{AC2} = -j\omega C_1 R_f V_s$$

This voltage will also be converted into dc signal and applied to the A/D converter as signal voltage V_{sig}. Thus, the ratio r of V_{AC2} and V_{AC1} is

$$r = \frac{V_{AC2}}{V_{AC1}} = \frac{C_1}{C_1 + C_2}$$

The ratiometric change *x* of the differential capacitance of the accelerometer can be given by [11]

$$x = \frac{C_1 - C_2}{C_1 + C_2} = \frac{2C_1}{C_1 + C_2} - 1 = 2r - 1$$

By measuring the ratio r of V_{AC2} and V_{AC1}, we can know the ratiometric differential capacitance change x. This will in turn be converted into a digital representation by the A/D converter as the final output.

2.2. BISR Control Circuit Design

To support the BISR MEMS structure, the BISR control circuit must be designed to electronically isolate a faulty module and replace it with a good redundant module. This replacement is implemented by a group of analog switches made of transmission gates (TGs) as shown in Figure 3. Here, all six identical modules of the BISR comb accelerometer are labeled with *a~f*. For each module, the pair of differential capacitances are labeled with 1 and 2 respectively. For example, C_{e1} and C_{e2} stand for the differential capacitance pair of module *e*. The selection signals for all six transmission gates are T_a to T_f respectively. The movable portions of all modules are connected together through common anchors, and are directly connected to the OPAMP A1 for signal sensing. The left fixed fingers of each module are connected to the voltage source V_s through an analog switch. The right fixed fingers of each module are connected to the common node of two NMOS transistors. The *on* and *off* states of each transmission gate decide whether the corresponding module is connected or separated from the main device. If the TG is *on*, the voltage source Vs is applied to the working capacitances of the module. The resulted current flowing through the capacitances of the module contributes to the total current sensed by the

OPAMP *A1*. Hence, the module is electronically connected into the main device. If the TG is *off*, the voltage V_s cannot be applied to the working capacitances of the module. Thus, the current flowing through the module is zero, so the module does not contribute any current to the OPAMP input. In this way, the module is electronically separated from the main device. The BISR control circuit sets the selection signals to their appropriate states according to the BIST result of each corresponding module.

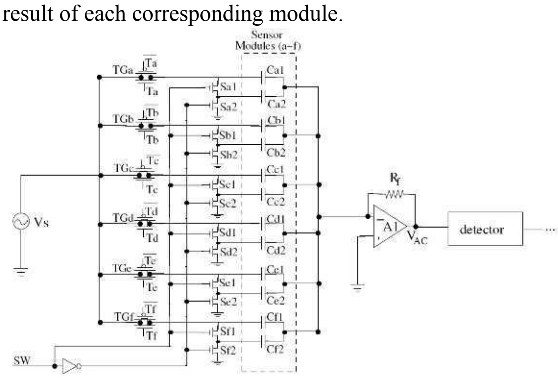


Figure 3. Switching circuit for redundancy repair.

When the device enters the BISR mode, it will first perform the BIST process [12] for each individual module to determine whether it is good or faulty. According to the BIST result, the BISR controller will set the selection signals T_a~T_f to their appropriate values. If any of the four working modules (initially assigned by default) becomes faulty, the BISR controller will set the corresponding TG selection signal of the faulty module to 0 to separate it from the main device. Simultaneously, the BISR controller changes the selection signal of a good redundant module from 0 to 1 to activate it as a working module, and connect it to the main device electronically. After the BISR process is completed, the MEMS device enters the normal mode. During the normal mode, four of the selection signals T_a~T_f are set to "high" and the other two are set to "low". That is, four out of the six modules (restructured by the BISR controller) will be connected as the main device, while the other two modules are separated as either redundant or defective ones.

III. Parasitic Analysis of BISR Circuitry

3.1. Parasitic Capacitance Analysis

Due to the tiny size of MEMS devices, their working capacitances are generally very small (in the range of 1pF or below). The parasitic capacitances of a MEMS device can easily go beyond this value. Hence, it is necessary to find out the influence of the parasitic capacitances, and a parasitic capacitance model needs to be developed for the analysis.

Taking module *a* in Figure 3 as an example, a model taking into account the parasitic capacitances is shown in Figure 4 where all major parasitic capacitances are extracted by ANSYS. Note that C_{a1} and C_{a2} are the differential working capacitances of module *a*, while C_{a10}, C_{a20} and C_{a30} are the parasitic capacitances between the left fixed fingers/right fixed fingers/movable fingers and the ground, respectively. Further, C_{a12} is the parasitic capacitance between the left fixed fingers and right fixed fingers. If the transmission gate TG and both NMOS transistors are treated as ideal switches, in phase 1, S1 is on and S2 is off, and the circuit can be simplified as Figure 5(a). In phase 2, S1 is off and S2 is on, and the circuit can be simplified as shown in Figure 5(b).

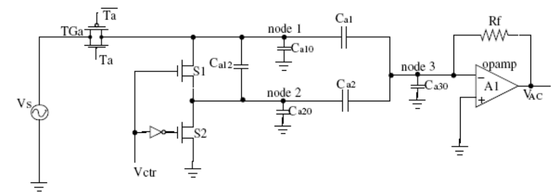
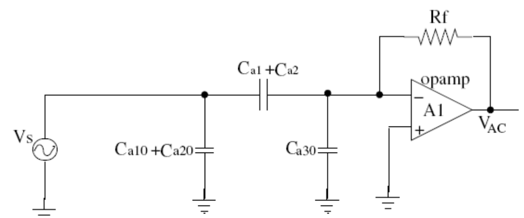
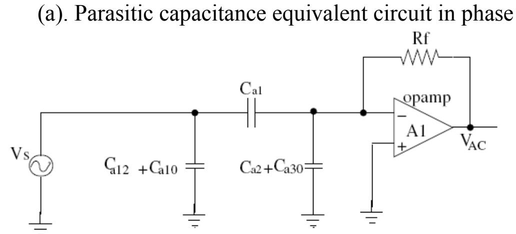


Figure 4. Parasitic capacitance analysis for one BISR module.



(a). Parasitic capacitance equivalent circuit in phase 1



(b). Parasitic capacitance equivalent circuit in phase 2

Figure 5. Simplified parasitic analysis circuit in two phases.

Assume the open-loop gain of the OPAMP *A1* as *A*. The output voltage V_{AC} in phase 1 can be expressed as

$$V_{AC1} = \frac{-V_{in}R_f\omega(C_{a1}+C_{a2})}{1+\dfrac{1+R_f\omega(C_{a1}+C_{a2}+C_{a30})}{A}}$$

The output voltage V_{AC} in phase 2 can be expressed as

$$V_{AC2} = \frac{-V_{in}R_f\omega C_{a1}}{1+\dfrac{1+R_f\omega(C_{a1}+C_{a2}+C_{a30})}{A}}$$

From the above analysis, we can see that only parasitic capacitance C_{a30} has influence on the output voltage, while other parasitic capacitances (e.g., C_{a10}, C_{a20}, C_{a12}) will not affect the output voltage. Since the open-loop gain *A* of an OPAMP is generally very large ($A \geq 10^5$), the denominators of the above two equations are approximately 1. Hence, the influence of parasitic capacitances will not be a problem for signal detection. In reality, the transmission gate and both NMOS transistors cannot be ideal. Although their off-resistances can be treated as infinity, their on-resistance are about 16kΩ and 32kΩ in our model.

In this case, the equivalent circuits of both phases for parasitic effects can be derived. Here, mathematical expressions of the output voltage for both phases cannot be directly accessible, and HSPICE simulations have been used for analysis. The transmission gates and NMOS switches are designed with Magic and extracted for HSPICE simulation. The device parasitic capacitances are extracted through ANSYS. Simulation results showed that the BISR device maintained the same sensitivity (V_{AC2}=6.83mV/g) with and without parasitic capacitances being considered. For example, even when we have C_{a1}=0.15pF, C_{a2}=0.05pF, $C_{a10}=C_{a20}=C_{a30}$=2pF, the circuit can still work properly for signal detection.

3.2. Analog Multiplexers and Signal Strength Analysis

In the BISR design, transmission gates are used to determine whether a module is connected to or separated from the main device. The output signal of a MEMS device is generally weak due to its tiny size. It is necessary to analyze the possibility of signal degradation caused by the added transmission gates. In order to ensure that the signal strength is not weakened, it is important to avoid passing the MEMS signals through transmission gates. In a MEMS accelerometer, generally, the signal is sensed from the movable portion. Thus, in our BISR design, the movable portions of all modules are intentionally connected together through the common anchors between neighboring modules. They are in turn directly connected to the OPAMP without bypassing through transmission gates. This ensures that the small current signal sensed from the movable fingers will not be degraded. The transmission gates are inserted between the signal source and the left fixed fingers of each module to determine the selection of the modules, as shown in Figure 3. The internal resistance of the voltage source is very small when compared with the on-resistance of the transmission gate R_{on}, which is around 16kΩ. Assume the total capacitance of the MEMS device is about 1pF, and the frequency *f* of voltage source V_s is 1MHz. The impedance of the MEMS device capacitance can be estimated as

$$Z_c = \frac{1}{2\pi f C} = 159.2k\Omega >> R_{on}$$

Thus, the on-resistance of each transmission gate can be ignored when compared with the impedance of the MEMS device capacitance. The signal source V_s can pass the transmission gate without significant degradation.

In summary, all movable plates are jointed together and connected to the signal sensing circuit directly, while a transmission gate is inserted between the left fixed fingers of each module and the voltage source.

In this way, the insertion of transmission gates will not degrade the signal strength sensed by the moving fingers. This has been verified using HSPICE simulation. All transmission gates are designed using *Magic* and the layout is extracted to *HSPICE* for simulation. This ensures that the parasitic effects from transmission gates are all considered. Simulation results show that the BISR device has nearly the same signal strength as the non-BISR device.

3.3. Defective Module Isolation and Load Effect Analysis

The separation of a faulty module from the main device is accomplished through the transmission gates inserted between the left fixed fingers and the voltage sources. As discussed, the movable portions of all modules are physically connected together. When a faulty module is separated out of the main device by turning off the corresponding transmission gate, its movable plate is still connected to the main device. Whether this will cause any load effect to the sensing circuit must be analyzed. According to the working principle of the BISR MEMS structure, the current flowing through the working capacitances in each module will be summed at the input of the OPAMP. If the transmission gate of the faulty module is turned off, there will be no current flowing through the working capacitances of that module. Hence, the contribution of current flow from the faulty module is zero. In this way, the capacitive load of a defective module will not affect the signal strength of the accelerometer. This has been verified with HSPICE simulation. For example, we ever changed the capacitance of a faulty module to different values (very large or very small), and the output signal remains the same as before. This proves that the capacitance of a faulty module will not cause any load effect to the sensing circuit. Even if the capacitance of the faulty module is changed to extremely large or small value due to defects, the output signal of the sensing circuit is still not affected. Further, we introduced a bridging defect by connecting a pair of movable and fixed fingers with a small resistor (0.001Ω), and the output signal is not affected either. This is because the fixed fingers of each faulty module are connected to the insulation layer instead of the silicon substrate. In case of bridging defect between movable and fixed fingers in a faulty module, the fixed fingers of the faulty module just share the same voltage level as the movable plates, and the movable fingers will not be shorted to ground. This demonstrates that the BISR technique is effective to bridging defect between the movable/fixed fingers. Actually the BISR design is effective to almost any kind of local defect which falls into an individual module. However, for global defects which occur to every module, the redundancy repair cannot be applied since there is no working module.

3.4. BIST Circuit for BISR

In order to implement the built-in self-repair technique of a MEMS device, the device must first perform built-in self-test for each individual module to decide whether it is faulty. The BIST result of each module will be fed into the BISR control circuit. Based upon this information, the BISR control circuit will virtually disconnect the faulty module and replace it with a good redundant module. In this way, MEMS BIST is the prerequisite for MEMS BISR.

Without an efficient and robust BIST solution, MEMS BISR cannot be realized because the control circuit cannot know which module is good and which module is faulty. A dual-mode BIST scheme has been proposed in [12].

Circuit support for the BIST method, especially the voltage biasing scheme, has also been presented in [12]. The proposed dual-mode BIST can serve as an effective BIST solution for the BISR of capacitive MEMS devices.

3.5. Design and Simulation of BISR Accelerometer

The geometry parameters of the BISR comb accelerometer with m=2 and n=4 are listed in Table 1, using a set of design rules comparable to ADXL accelerometers [2]. For comparison, a none-BISR accelerometer with the same number of capacitance groups (as that at the main device of the BISR accelerometer) is also designed. The geometry parameters of the non-BISR accelerometer are also listed in the same table. The simulation results for the performance of both BISR and none-BISR accelerometers are shown in Table 2. From Table 2, we can see that, by narrowing the beam width, the sensitivity loss of the BISR accelerometer due to device modularization can be fully compensated. The BISR accelerometer shows approximately the same displacement sensitivity as that of the none-BISR accelerometer.

Table 1. Design of BISR/non-BISR accelerometers.

Design Parameters	BISR device	non-BISR Device
device area (μm^2)	1500×900	980×900
thickness t (μm)	6	6
no. of capacitance groups	20×6	80
capacitance gap d_0(μm)	2	2
beam width W_b(μm)	2	3.2
beam length L_b(μm)	300	300
mass width W_m(μm)	200	200
mass length L_m(μm)	220×6	880
finger width W_f(μm)	4	4
finger length L_f (μm)	200	200

Table 2. Performances of BISR/non-BISR accelerometers.

Performance	BISR device	non-BISR Device
Sensing mass M_s (μg)	0.84×4=3.36	3.36
Capacitance C_0 (pF)	0.103×4=0.41	0.41
Sensitivity S_d (nm/g)	6.8	6.64
Sensitivity S_c (fF/g)	0.7×4=2.8	2.74
Spring constant k_m *(N/m)*	1.21×4=4.84	4.95
Frequency f_0 *(kHz)*	6.05	6.12

Given a non-BISR accelerometer design, its sensitivity and resonant frequency can be determined. As discussed before, in order to maintain the sensitivity, several alternatives are available. Assume the sensitivity loss is compensated by shrinking the beam width and enlarging the mass width simultaneously. How much should these two parameters be adjusted? The curves of displacement sensitivity S_d v.s. different beam/mass width values are drawn with MathCAD as shown in Figure 6.

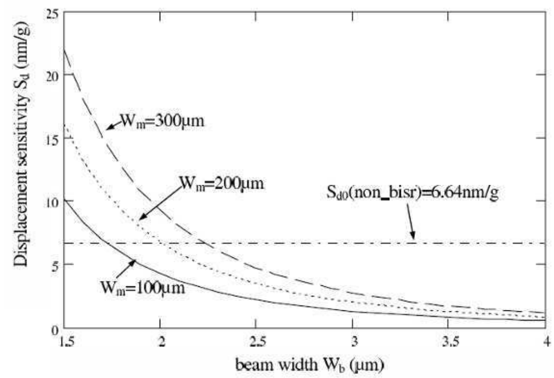


Figure 6. Sensitivity of BISR accelerometer design.

The sensitivity of the non-BISR design is also shown in the same figure as a horizontal line. The cross points between the sensitivity curves and the horizontal line suggest the possible solutions for the BISR design, with the same sensitivity as that of the non-BISR design. The frequency compensation can also be performed in a similar way (Figure 7).

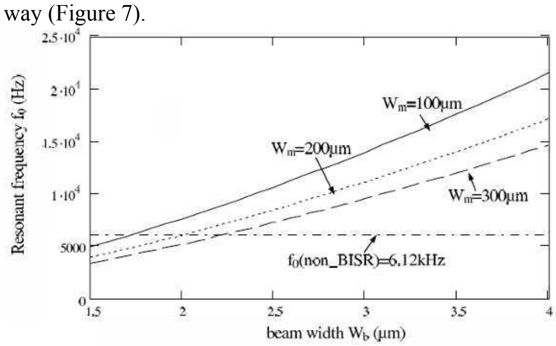


Figure 7. Frequency of BISR accelerometer design.

As shown in Figure 7, the displacement sensitivity of the non-BISR accelerometer is demonstrated as the horizontal line S_d=8.1nm/g.

The beam/mass width of the non-BISR design are 3.2μm and 200μm separately, as given in Table 1. If the beam width of the BISR design is shrunk to W_b=2μm, then displacement sensitivity will be approximately the same as that of the non-BISR device.

ANSYS simulation results for the displacement sensitivities of both non-BISR and BISR accelerometers in response to acceleration from 0 to 50g are shown in Figure 8. For the non-BISR accelerometer, the beam width W_b is 3.2μm. For the BISR MEMS accelerometer, the beam width is shrunk to 2μm. According to Figure 8, the sensitivity of the non-BISR accelerometer (expressed as the voltage output of OPAMP A_1 in phase 2, i.e., V_{AC2}) is 6.68mV/g, while the sensitivity of the BISR accelerometer after compensation is 6.83mV/g. The sensitivity of the BISR device is about the same as that of the non-BISR device.

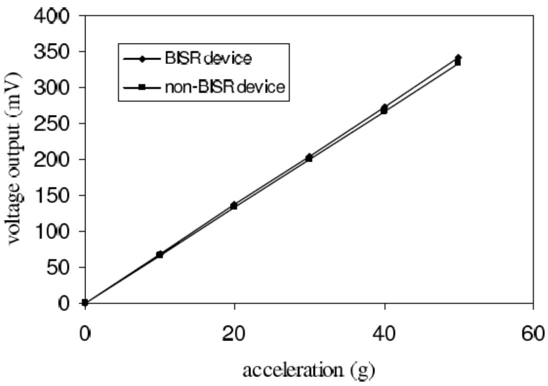


Figure 8. ANSYS simulation results for sensitivity of BISR/non-BISR accelerometers.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, the control circuitry for BISR MEMS comb accelerometer is proposed. A transmission gate (analog switch) is inserted between each individual module and the differential capacitance sensing circuit. Whenever a module is found to be faulty during the built-in self-test, the BISR control circuitry will switch off the transmission gate for the faulty module and turn on the transmission gate of a good redundant module. In this way, the whole device is self-repaired. The parasitic capacitance model of the BISR control circuitry is analyzed. The analysis results show that the parasitic capacitance will not affect the proper operation of the BISR control circuitry. Furthermore, the signal strength will not be degraded due to the insertion of analog multiplexers. The control circuitry can effectively isolate the faulty module of the BISR MEMS comb accelerometer.

In the future work, we will continue to develop the built-in self-test circuitry for the BISR MEMS comb accelerometers. Furthermore, there may be slight differences (e.g. sensitivity, working capacitance, parasitic effects) among the modules due to process variation. As a result, when a faulty module is replaced by a good redundant module, the whole device needs to be re-calibrated. In the future, a built-in self-calibration circuitry will also be developed so that the BISR device will maintain the same sensitivity after redundancy repair.

REFERENCES

[1] Navid Yazdi, Farrokh Ayazi, and Khalil Najafi, “Micromachined Inertial Sensors”, *Proceedings of the IEEE*, Vol. 86, No. 8, Aug. 1998, pp. 1640-1659.

[2] W. Kuehnel, and S. Sherman, ”A Surface Micromachined Silicon Accelerometer with On-chip Detection Circuitry,” *Sensors and Actuators A: Physical*, Vol. 45, pp. 7-16, 1994.

[3] “ADXL150/ADXL250: ±5g to ±50g, Low Noise, Low Power, Single/ Dual Axis iMEMS Accelerometers”, *ADXL150/ADXL250 datasheet*, URL: http://www.analog.com.

[4] B. Stark (editor), “MEMS Reliability Assurance Guidelines for Space Applications”, Jet Propulsion Laboratory Publication 99-1, Pasadena, USA, Jan. 1999.

[5] X. Xiong, Y. Wu, and W. Jone, "Design and analysis of self-repairable MEMS accelerometer," *Proceedings of the 20th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT'05)*, Monterey, CA, USA, pp. 21-29, Oct. 3-5, 2005.

[6] X. Xiong, Y. Wu, and W. Jone, “Reliability Analysis of Self-Repairable MEMS Accelerometer”, *Proc. of 21th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT'06)*, pp. 236-244, Arlington, VA, USA, Oct. 2006.

[7] S. Ogawa, Y. Oisugi, K. Mochizuki, and K. Watanabe, ”A Switched-Capacitor Interface for Differential Capacitance Transducers”, *IEEE Transactions on Instrumentation and Measurement*, Vol. 50, Issue 5, pp. 1296-1301, 2001.

[8] S. Chatzandroulis, D. Tsoukalns, ”Capacitance to Frequency Converter Suitable for Sensor Applications Using Telemetry”, *Analog Integrated Circuits and Signal Processing*, Vol. 27, No. 1-2, pp. 31-38, Apr. 2001.

[9] J. C. Lotters, W. Olthuis, P. H. Veltink, and P. A. Bergveld, ”A sensitive differential capacitance to voltage converter for sensor applications”, *IEEE Transaction on Instrumentation and Measurement*, Vol. 48, Issue 1, pp. 89-96, 1999.

[10] K. Mochizuki, K. Watanabe, T. Masuda, and M. Katsura, ”A Relaxation-Oscillator-Based Interface for High-Accuracy Ratiometric Signal Processing of Differential-Capacitance Transducers”, *IEEE Transactions on Instrumentation and Measurement*, Vol. 47, Issue 1, pp. 11-15, 1998.

[11] K. Mochizuki, T. Masuda, and K. Watanabe, “An Interface Circuit for High-Accuracy Signal Processing of Differential-Capacitance Transducers”, *IEEE Transactions on Instrumentation and Measurement*, Vol. 47, No. 4, pp. 823-827, Aug. 1998.

[12] X. Xiong, Y. Wu, and W. Jone, “A Dual-Mode Built-In Self-Test Technique for Capacitive MEMS Devices”, *IEEE Transactions on Instrumentation and Measurement,* Vol. 54, No. 5, pp. 1739-1750, Oct. 2005.

The Learning Process Management in E-learning Environment in the Technology School "Electronic Systems" Associated with the Technical University of Sofia

Ludmilla Yordanova Stoyanova

Technology School "Electronic System" associated with the Technical University of Sofia,
Mladost 1, 1750 Sofia, Bulgaria

Abstract-**The goal of this paper is to analyze the possibilities for learning process management in the e-learning environment of the Technology School "Electronic Systems" associated with the Technical University of Sofia. In the paper interactions between student, lecturer and content are described. On that base the test results received from the system are proposed to be analyzed in two ways in order to help the lecturer taking learning process management decisions, concerning the further students' activities – individual or in group.**

I. INTRODUCTION

The modern education requires the usage of an e–learning system on a Web platform. The implementation of such system has a wide range of advantages [3] in the educational process for the students and for the lecturers. The e-learning environment in the Technology School "Electronic Systems" (TUES) has been developed, implemented and used for four years and during this period of time the platform and the functionality as well have changed.

The recently used information system of TUES has the following basic modules:

- official part - consists of home page, students' admission module, news etc.;
- management of different categories of users - students, ex-students, teachers and administrators;
- e-learning system – the purpose of this module remains making educational process more interesting and attractive to students as well as facilitating the work of teachers related to process of teaching and evaluating the results shown by students [4];
- software and hardware resources [1].

The e-learning system in the Technology School "Electronic Systems" is not being used to perform distance education. Its purpose is to support the face to face educational process and to improve its management. It has two sub modules – the student's module and the teacher's module.

The student's and the teacher's modules realize faculty and students communication, sending messages, information and curricula exchange, discussions forum, subject matter and course work. The student's module has been designed to assure the ability of the students to reach to the lecturers' information and use the e-learning system module – reading materials, testing knowledge and skills etc.

The communication between students and lecturers in the face to face learning process has been influenced from the implementation of the e-learning system.

II. INTERACTIONS IN THE E-LEARNING SYSTEM

The usage of the e-learning system gives additional possibilities for interactions between student, teacher and content. The interactions between lecturer and student are not only in the classroom but also in the e-learning environment as shown on fig.1. This makes the communication easier and more flexible.

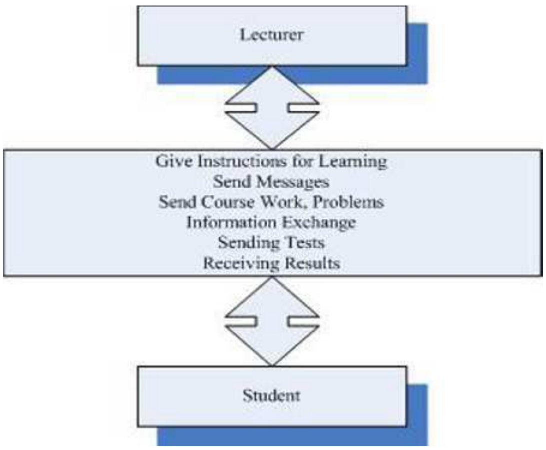


Fig.1. Interactions between lecturer and student in the e-learning system

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_50, © Springer Science+Business Media B.V. 2010

The teacher's part has been developed for the lecturers in order to give them ability to upload information - lectures, tests and other materials. They are also able to create, modify and publish multimedia educational materials in the e-learning module of the information system.

The e-learning system contains tools for creation of multimedia educational materials for different disciplines in TUES. It maintains tools for self-education and test system for knowledge and skills assessment (fig. 2).

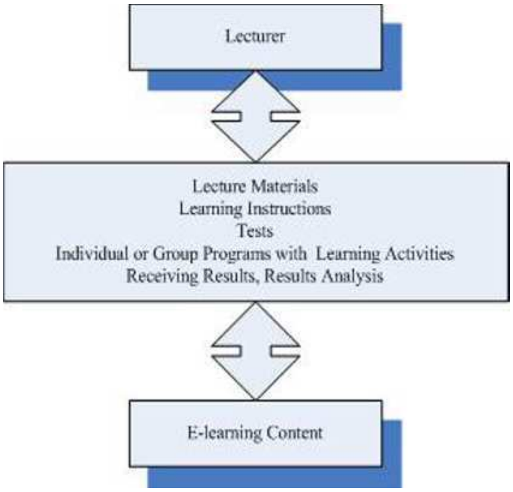


Fig.2. Interactions between lecturer and content in the e-learning system

The lecturer plans, designs and creates the educational content in the e-learning environment:

- lecture materials, exercises, problems and different kinds of tests etc., creating the hierarchical structure of the learning content ;
- type of tests evaluation and the time of test fulfillment;
- improvement or change in a definite way of the created learning materials on the base of the control and evaluation of students results;
- programs with further learning activities for each student or for a group of students.

The student works individually or collaborating with the others in groups in the e-learning environment – he reads the materials, fulfils the sent exercises, solves problems and makes tests or he performs course work sometimes together with other students (fig. 3). He can see his results and fulfill additional self education and self control testing. The usage of the created from the lecturer content in the e-learning system is very popular in TUES.

On the base of lecturer's work in the e-learning environment he improves his interactions with the students. The communication between the lecturer and the student is realized in the:

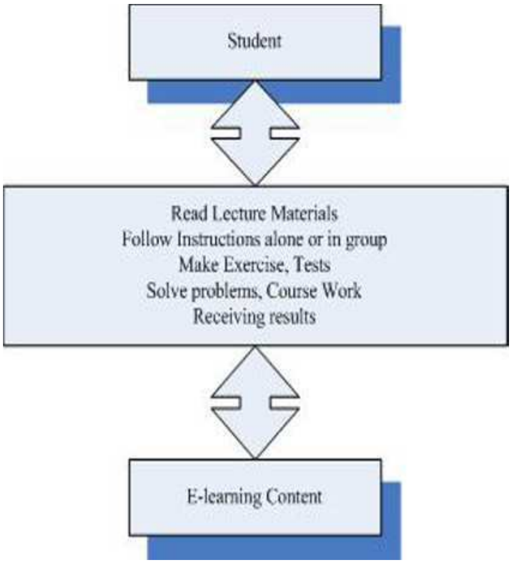


Fig.3. Interactions between student and content in the e-learning system

- e-learning environment – upload and download of materials, exchanging messages, doing tests, receiving test results, sending and receiving mails. There are also possibilities for discussions forums, but it is rarely used in TUES;
- face to face learning process in the class room with all ordinary forms of communication.

The lecture materials in the classroom are presented for a whole group of students while in the e-learning environment the material could be individually intended. The consultation for the students could be realized using the different possibilities of the information system (listed above) individually or for a group of students. The work in the classroom could be widely affected by the results achieved in the e-learning environment and can give result in individual tasks for the student.

The communication between student and lecturer (fig. 1) can be realized by:

- sending messages;
- mailing problems, course work, etc.;
- information exchange;
- sending tests and receiving results;
- discussions forum (very seldom used);
- face to face instructional education in the classroom.

The lecturer communication with the student is face to face in the classroom and over the e-learning environment and has the purpose to help the student in his learning process.

In fact the student consults the lecturer about his problems or difficulties in the education mainly in the classroom education. The proposed in the information system possibilities to communicate the teacher in the form of consulting are rarely used.

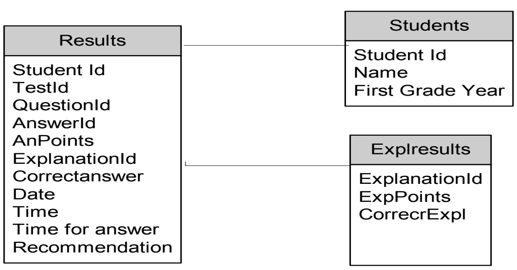
The sent course work, problems etc. from the student using the e-learning environment are not to be valuated from the system, but manually from the lecturer.

All data that the e-learning system receives as feedback that can be used for further learning process management by the system itself are the test results. On the base of these results the system can help in preparation of learning instructions for each student or for a group of students.

III. LEARNING PROCESS MANAGEMENT

The learning process management is realized on the base of the set goals and the achieved results from the students. The testing of students in groups gives many advantages for the lecturer. Usage of test results analysis for this group would help the lecturer in his further activities in class or in e-learning system.

The e-learning system supports the lecturer with additional information about the achieved results from the students for his further management decisions concerning the learning process. The saved data from the fulfilled tests and exams in the database structure are shown on fig. 4.



Fi

Fig. 4 Test results accumulation in the database structure

In the Students table is the data for student's identification. In the Results table (fig. 2) the identification of the test, question, answer, points etc. and the connection to the Students table are kept. The table Explresults has been used for exact calculation of the points for each given answer of the student [2].

After the fulfillment of a definite test it would be useful to know the distribution of percentage of shown knowledge and skills. The percentage of shown knowledge and skills are grouped in the well known in Bulgaria official groups for defining the marks of the students and listed bellow:

0 … 50% - mark 2 (poor);
51 … 70% - mark 3;
71 … 81% - mark 4;
71 … 91% - mark 5;
91 … 100% - mark 6(excellent);

TABLE 1. DISTRIBUTION OF SHOWN KNOWLEDGE AND SKILLS IN A DEFINITE TEST OR GROUP OF TESTS

KNOWLEDGE SKILLS	0 - 50%	50 - 70%	70 - 80%	80 - 90%	90 - 100%
STUDENTS RESULTS	N1	N2	N3	N4	N5

The distribution of marks for a test or a group of tests is important for the lecturer for it shows how the set goals are realised.

In table 1 is presented the table for such analysis. The numbers N1, N2,…, N5 represent the calculated percentage of shown knowledge and skills for a test or for a group of tests. They are calculated as shown in (1).

$$N\,x = Mx / M \,.\, 100\,\% \qquad (1)$$

In (1) the Mx is the number of students with test result within the appropriate range of shown knowledge and skills, M is the number of all tested students (x shows the number of the range which usually is from 1 up to 5 but can be changed).

This analysis helps the teacher to take the decision for additional training of a group of students and the type of training – in the e-learning system or in the classroom. But there are cases in which the system can help the lecturer in the analysis. In table 2 an example of fulfilled test for the distribution of shown knowledge is shown. The test has been done by 52 students. The decision for additional training can be helped by giving appropriate directions to the lecturer from the e-learning system. The preparation of these directions can be done by the lecturer himself at the beginning of a course. He would be able to define the minimal (or maximal) percentage of students mainly in the first and last group (poor marks and excellent marks) and the corresponding activities to solve the problem or to go on with the curricula.

TABLE 2. EXAMPLE OF DISTRIBUTION OF SHOWN KNOWLEDGE AND SKILLS IN A DEFINITE TEST

KNOWLEDGE SKILLS	0 - 50%	50 - 70%	70 - 80%	80 - 90%	90 - 100%
STUDENTS RESULTS	12	61	13	14	0

All this could be followed by a deeper analysis – for the lecturer would be useful to know the distribution of given answers for every question in the test for the whole group of students. In table 3 is presented the distribution of given answers for a definite test.

TABLE 3 DISTRIBUTION OF GIVEN ANSWERS IN A DEFINITE TEST

QUESTIONID	CORRECT ANSWER (%)	WRONGANSWER[1] (%)	… %	WRONG ANSWER [N] (%)
1	50	10		10
2	0	20		80
3	98	0		0
.	.	.		.

In this way not only the correct answers are taken into consideration but the proposed wrong answers too. The proposed wrong answers given by the students present the most typical mistakes done by the students. This suggests description not only of correct answers for a question but also the proposed wrong answers. The work done would give the option for the lecturer to put in advance into a test recommendations for further learning for a student or groups of students without his personal engagement for such analysis.

The high percentage of wrong answers to a group of questions is a condition for taking decision for additional explanations and training of the group of students in the lecture material represented by these questions – giving advice to revise a lecture or group of lectures, or even a definite problem.

IV. CONCLUSION

The e-learning environment allows the lecturer to plan for each student or for a group of students individually the access to a definite kind and type of learning resources.

The test results should be analyzed from the distribution of the percentage of shown knowledge and skills point of view. That planning could be done in advance by the lecturer for the typical cases. It could be realized on the base of analysis of test results.

This distribution allows the teacher to decide in advance how to continue his further activities – proceed with the learning materials, fulfill additional training for a definite student or for a group of students in the e-learning system or face to face in the classroom. The deeper analysis of the test results, that is the distribution of given answers (correct and wrong answers), allows the lecturer to define the difficulties in the learning of a student or of a group of students and prepare the instructions for typical cases.

These analyses inside the e-learning system help the lecturer with his further planning of his own activities and of the activities of the students as well.

V. REFERENCES

[1] Stoyanova L., Stefanova S, Chorbadjiev L., *Information system of the Technology school "Electronic systems" associated with the Technical University of Sofia based on CMS*, The Int. Conference ELECTRONICS'2005, Sozopol, 2005

[2] Stoyanova L., Test results assessment and analysis for the Information system of the Technology school "Electronic systems" associated with the Technical University of Sofia, . The Int. Conference ELECTRONICS'2005, Sozopol, 2005

[3] Graham Attwell, Lone Dirckinck-Holmfeld, Peter Fabian, Andrea Kárpáti and Peter Littig "E-Learning in Europe – Results and Recommendations", Thematic Monitoring under the Leonardo da Vinci Programme, http://www.elearningeuropa.info/, March 2004

[4] Stela Stefanova, Ludmila Stoyanova, Sergey Petkov, Petar Todorov, User management in the e-learning information system of the Technology school "Electronic systems" associated with the Technical University of Sofia - unpublished

Taxonomy of E-courses[1]

Leszek Rudak, Dorota Sidor
Centre for Open Multimedia Education University of Warsaw
Ul. Smyczkowa 11a/2, 01-678 Warszawa, Poland
{l.rudak, dorota.sidor}@uw.edu.pl

Abstract: **Three attributes of online courses were described in the article: the communication, the structure and the organization, which values characterize every course. All acceptable values of these parameters where outlined as well as a usefulness of this taxonomy for classification of vocational courses was examined.**

I. Introduction

In designing a course that is going to be conducted through the Internet, one of the most important decisions is about the choice of course type. While designing we have to know whether we want to have full-time, stationary classes (if so, which content is going to be presented in this way, and which on-line), what will be the role of a teacher in the course, are there any expected discussions between the participants, group work, open tasks, what pathways will student have to go through with the content of the classes, etc. A choice of particular course type (model) gives us the answer to most questions.

In literature we can meet different approaches to the problem of indicating types of on-line courses. In [17], analyzing the way of transmission from traditional teaching to e-teaching at American universities, the author distinguished five types of courses: *supplemental* (supplementing a stationary course with materials available on the Internet), *replacement* (all classes are replaced with on-line), *emporium* (part of traditional classes is replaced[1] with on-line classes, but every student can at every time – duties are at least for 50 hours per week – consult with the academic teacher at the university), *fully on-line* (all classes are conducted with the usage of the computer and the Internet, they are conducted by teachers who are available on-line), *buffet* (free choice of the way of education allows student to have every classes conducted in a different way – on-line, thanks to materials from the Internet, thanks to direct consultation with the teacher, or thanks to stationary classes). The criterion for this division is the number of stationary classes that are conducted and the on-line activity of the person conducting. Unfortunately it does not give us clues which support designing a course because the categories presented are very broad and include a lot of different types of on-line courses. For instance courses *fully on-line* include almost all types of e-courses, whereas *supplemental* category nowadays include most traditional classes – a lot of university teachers place supplementary materials for students on the web (their own notes, tasks to do at home, clues for homework, etc).

In [8] the basis for course division was the proportion of interaction (through the Internet) between student and teacher in comparison to time spent by the student on the whole course. There were three types of distinguished courses: *content + support* (materials are available on the Internet; student receives a little of teacher's help – approximately 20% of time spent by the student on the course), *wrap around* (on the Internet there are materials available which supplement the standard materials; the time of interaction between the student and the teacher and other students equals approximately 50% the time spent on the course), *integrated* (the interaction between the student and the teacher and other student is the main part – discussions, solving problems together, fulfilling group tasks, etc). This division is not sufficiently precise so that it could help in designing courses. In particular, the *wrap around* category is so broad that it can include most courses conducted on the e-learning platform where discussion forums are used.

Another way of indicating types of courses was used in [3]. The authors pointed activity of the student during classes as the main indicator of course type. They distinguished six types of courses: *Socratic* (based on discussion between students and teacher performed through the Internet), *person conducting / student* (the teacher on-line shows the direction of student's work), *group work* (students realize, with the usage of computer and Internet communication, common project), *small groups* (students fulfill tasks by working in small groups), *instructor / student* (course is an educational computer program that instructs student through required material) and *independent studies* (students has only access to materials available for him on the Internet). This division does not include hybrid teaching where part of classes is taught in a stationary way and the other part on-line.

In this article there is taxonomy of internet courses presented that is based on three attributes, their values may be easily defined for each course. On the other side, the values of those attributes may be defined also when there are known aspects such as educational goal, content that is going to be presented during the course and realization capabilities. Thanks to that, in the beginning phase of designing we can define the category of the course appropriate to the didactic assumptions. Attributes and their values result from the analysis of the didactic situation in remote (on-line) teaching with the usage of the Internet.

[1] Paper and research supported by the European Social Fund: "Developing of methodologies, programs and didactic materials to the distance education."

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_51,

II. Didactic situation

Every type of education consists of at least three elements: teacher, student and content of teaching. The only exception is self-teaching which means a situation, where somebody without any help of the other person acquires knowledge by studying or searching information in the data base [11]. It does not refer to using textbooks or computer or self-teaching, where the didactic process designed and included in the textbook or program has the role of the teacher. All elements of the didactic situation are in relations with each other, which is presented on the figure 1.

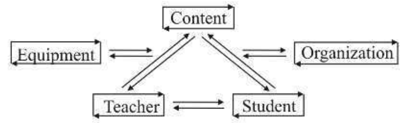


Fig. 1. Didactic situation from [11].

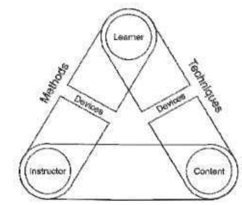


Fig. 2. Didactic situation of e-teaching from [4].

The most important of those: communication between the teacher and the student and also relation between the student and the content of the course, in e-teaching are interpreted appropriately as methods of social organization of students and techniques of giving/receiving knowledge and presenting skills. The connection between elements of didactic situation is maintained thanks to technical facilities/equipment which allow to communicate via Internet [4].

The diagram on figure 2 passes over the relation between the teacher and the content of the course. First of all it does not present that equipment and software also interfere in this connection – maybe even more than the other two. Often the way of presenting the material in a form of digital objects influences the presented content. In respect of the digital technology omnipresent in the e-course it is no use distinguishing equipment and software in the diagram, however there should be an interpretation of all relations presented.

The organization of communities in the process of learning is the mutual exchange of information and views between the teacher and students or only between students. That is why that relation should be interpreted not as a method of teaching but rather *communication*. Relation between the student and the content of the course is realized thanks to the access to the materials presented during the course. The order of making files available (texts, audio and video recordings, simulations, etc) which contain materials described in the course, is planned by the teacher and author of the course. This relation we interpret as the *structure* of the course. Finally the third relation – teacher and the content of the course refers to the factual aspect – teacher is responsible for what student is going to learn but also for the organization of the content accordingly to the course plan. In case of traditional teaching there is only one way of conducting classes: direct, sequential meetings between the teacher and the students. Whereas in e-teaching we can use different ways: beginning from on-line course (without any stationary meetings) to the classes almost fully stationary, supported only by on-line materials. So, the connection between the teacher and the content may be described as the *organization*, paying attention to the aspect of that relation described above. The diagram of didactic situation may be presented as in the figure 3.

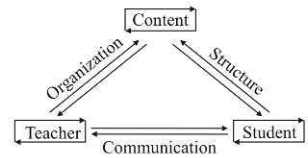


Fig. 3. Interpretation of relations in didactic situation

Teaching with the usage of web has to be very flexible. In theory we can distinguish six aspects of that flexibility: time, place, pace of teaching, communication medium, access to materials and syllabus of education [12]. Three of them: any time of studying, any place of studying and any, adjusted to student, pace of learning are inseparably connected with the paradigm of e-teaching. Whereas the three remaining are connected with relations resulting from a didactic situation. The medium provides opportunity for communication between participants of the didactic process, access to materials, defines the structure of the course, the syllabus defines organization of the education process.

Both discussed above ways of categorizing (the first one focused on didactic situation, the second one focused on flexibility of educational process) determine three features typical for every course that is conducted on-line. We can assume that communication, organization and structure of the course are the indicators of e-course category.

III. Communication

Communication between participants of the educational process conducted through the Internet uses transmission medium that is a computer (with an access to the Internet). In literature there are four types of such communication defined: *one-alone* (solitude), *one-to-one* (a dialogue), *one-to-many* (a lecture), *many-to-many* (a discussion) [16], [13]. However, this division does not illustrate all possible didactic situations – for instance a form of communication between two people depends on the party that initiates dialogue. If it is a student, then a conversation is in a form of consultation[2]: a student wants to get to know something new, eliminate doubts or ask

[2] In *one-to-one* form we assume that student and a teacher are the parties of the conversation; communication between two students we analyze only in a context of group work, but without limitation of the participants number.

for explanations. Whereas a conversation initiated by a teacher is mostly an open task or problem to be solved by the student. Each form of the communication has different role in e-course structure. Similarly, we should distinguish different variants of group communication (*many-to-may*): communication in a group of students without a teacher and communication with a teacher. The choice of one of the above types of communication results from didactic assumptions. To the extent that a discussion with the presence of all course participants is a natural method of education, transferred from a traditional lesson and improved with the technique of group discussion and discussion forum on the e-learning platform, then application of friend communication, which means in a group of students without a teacher, it can bring about some objections. However, we can present a lot of examples that such form of education has a big role, particularly for young people who learn thanks to the interaction with other people existing in the community. For adult people this form of communication is less important [10].

To sum up, deliberation about communication between participants of education process with the usage of computer and the Internet, we define presented types of communication by connecting directly a type of interaction with the e-course category.

Self-teaching is a term used to distinguish a course in which there is no communication between the participants (solitude, *one-alone*). Such classes are usually conducted as application delivered to the user on the external memory. We can also meet this type of course in a form of traditional, static websites.

Friend communication means a course, in which there is an opportunity of a discussion in a group of students (often in an asynchronic way, but the synchronic way is not excluded) without a teacher (*many-to-many*). An instructor may of course listen to the discussion, but as assumed without taking part in it.

Course of open tasks is a course that provides communication between the teacher and every student separately; the teacher is the party initiating *(one-to-one).* Such communication is usually performed with the usage of open tasks mechanism: a teacher gives a problem and waits for the solutions sent by the students, after that he sends his individual comment on every solution. In this way, there is no group communication.

Consultation is another form of communication between two people *(one-to-one)*. This time a student initiates the conversation who raises a question, doubts or comments directly towards the teacher. Further communication (explanations, additional questions, supplements) takes place only between those people. Usually this form is performed via e-mails (often being a service of e-learning platform).

Teacher's instructions indicate such type of e-course where teacher gives students additional content or organizational information without interaction from the other side (*one-to-many).* In this case we do not mean a form of interaction "learner-instructor", described in [10] or [9][3], but remote-controlled dynamic educational process which is not individualized. Here, we should underline that communication (*one-to-many)* makes sense only when teacher is the person giving students information.

Full communication means that in the course there are all forms of communication realized between participants (with the help of the computer), including discussions in a group with a teacher (*many-to-many*). Here, we should underline that it is important to use those mechanisms for didactic goals, not the possible opportunity for technical realization.

A form of communication in the presented taxonomy depends on the used methods of education, not the technical capabilities. That is why not every course placed on the e-learning platform will deserve to be called full communication, although the realization of discussion demand with the participation of all students and the teacher is easy to perform, for example with the help of discussion forum. Describing e-courses we have to use only one definition – which requires the least resources. For example, if the course is not *self-teaching* it means it contains some forms of communication, then it certainly uses the mechanism of mutual communication student-teacher. However, not every course is a *course of open tasks* because it may not use (or even it does not have an access to) mechanism of sending solutions to the tasks and comments as well as assessing them by the teacher. Such type of a course is a *consultation.*

IV. Organization

Education through the Internet, despite its advantages related to the breaking up with the unity of place, time and action of the didactic process, it also has disadvantages. It is not possible in on-line way to work with the factual object, we cannot carry out an experiment in a laboratory (particularly there is no possibility to experience smell, taste and other senses through the Internet) there is no possibility to come into interaction with other people with all features of a direct meeting (body language, influence of the environment, eye contact, etc) [1]. Those inconveniences may be solved by introducing to the e-course few or more meetings, lessons, lectures or workshops in a traditional form, in a traditional room, laboratory. Such form of classes is called *blended learning* (mix education, *hybrid education*).

Blended learning does not only solve problems which are not possible to solve by e-learning but also makes it easier to adjust educational forms to the demands of generation that was born in the age of computers and the Internet and hinders the process of dehumanization of education [7], [2].

The stationary part of *blended learning* may be devoted to different activities: introduction to the subject of the course, review and summary lectures, work in small groups, project presentations, demonstrations, laboratory and workshop activities, lectures visiting specialists, question and answer

[3] Based on the fact that the content of education is delivered by the teacher with the usage of materials placed in the course and the layout of the materials – such form is present in every type of e-course.

sessions or advanced factual discussions. However, stationary and on-line classes always have to be appropriately connected so that they are carefully planned as a whole [14]. In designing a hybrid course one should consider first of all the goal of learning and the schema of the course resulting from the goal. We can distinguish three very general models, different from each other in type of the main pathway of the stationary classes: oriented to the consolidation of skills, change of views and formulating the competencies [18]. Despite the fact that all those models have their own characteristics, then so general distinction does not allow to define operational goals and does not help in designing. However, when using the described features and the realization prepositions included in the papers [14], [18], [5] and [15] we can accept three very precise schemas of the hybrid courses differentiated on the basis of the content presented during stationary classes.

The first schema, *occasional course,* has a stationary part limited only to few meetings, usually in a the circle of all participants. They are devoted to all organizational issues but may also include introduction to the subjects and summaries of the content present in the course. Particularly, the inauguration meeting is very important ([14]) although it does not have to be obligatory in the course.

The second schema, *practical course*, where there is a stationary part devoted mainly to the laboratory or workshop classes, off-site activities, gathering of empiric information, etc. Of course, some elements of teaching theory may be present, but the main effort of students is concentrated on the acquiring of the practical skills.

The third schema is the opposite to the second one, because the stationary part of the *theoretical course* is devoted to the transmission of the theoretical knowledge through the lectures, demonstrations, and classes of seminar nature. Practical classes that is using knowledge in everyday situations is (almost) transferred to the Internet (such course schema seems to be good in teaching mathematics, economics and philosophy – in those subjects traditional lecture and the lecturer play a big role, and the practical classes are usually performed individually without necessity of direct participation of the teacher).

To take into consideration all possible cases in the division we have to introduce category of *on-line course* that is such course where there are no stationary meetings. We do not define a course without any classes on the Internet because we do not examine this form of education here.

V. Structure

Structure of internet course is defined by the access to the materials or fragments (course elements) into which the content is divided. There are a lot of opportunities to make resources available for students, but usually we can distinguish three solutions: line structure, topic and social [6].

Line structure is based on making resources available in defined time intervals, e.g. every Monday new materials are made available as a part of weekly content portion. Usually, student has an access to new materials without necessity to fulfill additional requirements. Very often every stage is completed with a test checking if the presented knowledge is acquired. *Time hierarchy* matches to this structure and this name will be used in the further part of this work.

Topic structure is a division of all course resources into groups which are related to the particular topics included in the syllabus. Materials in each group are placed in the particular order that illustrates the hierarchy of issues. Such structure may be present in two forms. In the first one, using materials inside the chosen subject requires to go through the "control points" – tests that check the level of mastering the material preceding this point. Too low mark does not allow to have an access to the further resources. Such type of course is called *Topic hierarchy*. The second form (and this one is described in the quoted above portal) does not have any conditions limitating opportunity for access to files. For this form *Independent studies* is the most appropriate name because it corresponds exactly to the model of e-course from the mentioned above taxonomy presented in [3].

Course organization around the activity, called social in the quoted above portal does not indicate particular nature of the course. It is drawing the attention to certain activating learning methods which may be successfully used in other types of courses, e.g. in the course with time hierarchy (the only structure that ensures the efficiency of the methods where more people take part). It is difficult to imagine a course with a construction that has only those activities.

To have a full image of possible internet course constructions there is one thing missing – the easiest structure: a collection of materials without any hierarchy. This form is often used in supporting stationary learning with resources placed on the Internet. Often direct meetings allow students to have appropriate clues on using the materials on the net so no structure of those resources is needed. Courses which contain manual how to use the equipment, fill in the questionnaires, apply new regulation and principles in an enterprise or office do not require any structure either. This type of course, that is one box with digital documents, is called a *Collection of materials.*

VI. Taxonomy

The analysis presented above has led to the definition of three attributes: *organization*, *communication* and *structure*, whose values indicate type of course. There are possible values of those attributes indicated. But we cannot claim that any three values (one for each attribute) indicate a course category because some of them are against each other. For example a *Collection of materials* as assumed does not take into consideration teacher's participation so it has to have communication *self-teaching* and organization *on-line course*. Other values for communication and organization are excluded.

Similarly, organization anticipating factual meetings (not occasional) indicates the course structure *Time hierarchy* because the efficiency of direct meetings may be provided only when all participants will be well prepared (master the advised range of knowledge or skills). Such conditions will not be provided by the *Collection of materials, Independent studies,* or even *Topic hierarchy* because in such conditions there is no control over the order of topics that is studied by an individual participant of the course. In the table presented below it has been shown which values of attributes may indicate type of e-course. It was accepted, except for assumption defined above, that it does not make sense to introduce factual discussion between students when there is no certainty that all participants had and access to the same materials, which excludes *Friend* and *Full communication* in case of *Topic hierarchy* and *Independent studies.* Moreover, we assume that both *Collection of materials* and *Independent studies* take place without any part of the teacher so they exclude blended learning. For all possible course categories see Table I.

VII. »EXPERIMENT«

Theoretical deliberations have led to indicating 31 different categories of courses. A question is important whether all those categories are needed: whether they are, or may be, any courses having those attributes. On the other hand we decided to verify our deliberations with the usage of some kind of experiment. We decided to check the usefulness of the presented taxonomy of course category on the module units of educational programs for professional qualifications based on the order from the Ministry of Education. The research was to check:

(1) Whether every (real) course may be described with the usage of three attributes: organization, communication and structure? (2) Whether all values of the attributes are necessary?

The essence of the ‘experiment’ was to analyze module units from the point of view of opportunities of their realization in a form of e-course (including possibility of blended learning), and then to catalogue such (hypothetical) courses with the usage of the given taxonomy.

A group of teachers and experts of e-learning from Centre for Open and Multimedia Education (COME) at University of Warsaw (A. Barbasiewicz, B. Michałowicz, M. Siarkiewicz and D. Sidor) verified 1400 module units. Having rejected professional internships (120 units) 1280 remaining were to be verified by the experts panel with the usage of accepted categories.

TABLE I
COURSE CATEGORIES

	Collection of materials				Time hierarchy				Topic hierarchy				Independent studies			
ORGANIZATION[a]	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Self-teaching	+				+				+				+			
Friend communication					+	+	+	+								
Open task course					+	+	+	+	+	+			+			
Teachers' instruction					+	+	+	+	+							
Consultation					+	+	+	+	+	+			+			
Full communication					+	+	+	+								

[a]A – on-line course; B – occasional course; C – theoretical course; D -- practical course.

VIII. RESULTS OF »EXPERIMENT«

A. Orgaznization

Part of the tested units cannot be conducted as internet courses. Among those units that may be performed as internet courses most require stationary meetings. It results from the specificity of professional education focusing on practical skills (Fig. 4).

The most popular model of organization was the one in which theoretical part was presented on line and the practical part in a stationary way. One more time, big part of practical knowledge, possible to be presented only directly, has determined the results of categorization (only four times there was a model ‘stationary part – theory’) (Fig. 5). Courses fully on line are the courses which formulate in the participants so called ‘soft’ skills or present abstract knowledge.

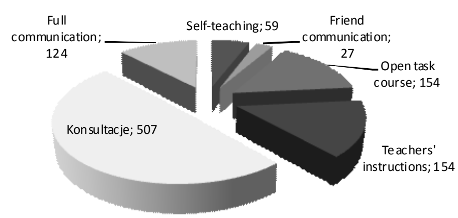


Fig. 6. Attribute: Communication

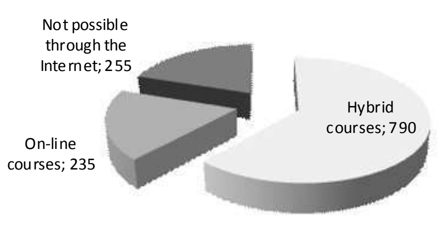


Fig. 4. Organization (all units).

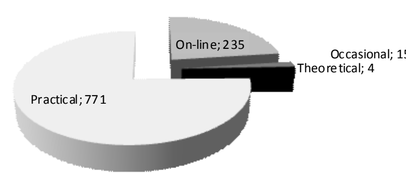


Fig. 5. Attribute: Organization

B. Communication

It results from our 'experiment' that majority of units require a certain form of communication with a teacher and/or other participants. Often present categories were: open tasks course, teacher's instructions and consultations (Fig. 6).

When analyzing the results of his attribute we hale to remember that some flexibility is allowed here – model of Communication is not given once for ever. In particular, it refers to all subjects where only communication teacher-student is taken into account. That is why the data referring to the open tasks course, instructions and consultations should be taken into account together. Those forms of communication are the most popular due to the limited number of courses where it is necessary to introduce generally available factual discussions. Because of the same thing there is few courses with full communication.

Assigning course to category *self-teaching* and *friend communication* depends mostly on the course structure – enable to communicate between students may take place only when a group of students begin classes at the same time.

C. Structure

The most popular model of presenting materials structure was time hierarchy course (Fig. 7). It results from a big number of hybrid courses which always have time structure. Necessity to organize stationary meetings nearly exclude opportunity to conduct subject course that from its definition does not have any terms, in which student would have to show that the theoretical knowledge was mastered by him (very often necessary while performing practical tasks). Few courses are realized in a form of set of materials and free course – majority of courses require introducing regular knowledge checking.

X. Conclusions

We may conclude that all (apart from internships and subjects only practical) module units may be (hypothetically) realized as internet courses characterized by presented in the taxonomy attributes and indicated values. Moreover, all values of the attributes were used to realize all units. The taxonomy offered seems to be complete and useful in categorizing e-courses which are to be used in professional education. We assume that this taxonomy is equally useful in categorizing university subjects because of their more abstract, and less practical nature which means bigger susceptibility to e-learning.

The taxonomy can be used in the course designing process. One can start with stating the educational goal and the suggested content of the prepared course. The content strongly indicates the organization of the course (if there is a need for face-to-face meetings: practical exercises in a workshop or meetings with special guests, etc.) while the goal depicts the minimal level of communication (e.g. if the goal is to prepare students to use particular equipment then no discussions are necessary. On the other hand, a good merchandiser has to practise his negotiation skills). Both, the

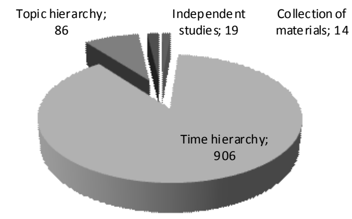


Fig. 7.Attribute: Structure

goal and the content point out the structure of the course (e.g. if the face-to-face meetings are necessary then the structure time hierarchy should be used). Thus we get the type of the course and its plan. So the suggested taxonomy simplified the first stage of the process of design e-courses.

References

[1] I. Bednarczyk and L. Rudak, "Podatność przedmiotów akademickich na e-nauczanie" in M. Dąbrowski, & M. Zając (Red.), *E-edukacja dla rozwoju społeczeństwa*, Warszawa: Fundacja Promocji i Akredytacji Kierunków Ekonomicznych, 2008, pp. 132-139.

[2] J. Bednarek and E. Lubina, *Kształcenie na odległość. Podstawy dydaktyki*, Warszawa: Wydawnictwa Naukowe PWN SA, 2008.

[3] G. L. Bedore, M. R. Bedore and G. L. Bedore, *Online education: The future is now*, Phoenix: Academic Research & Technologies, 1997.

[4] B. R. Burnham and B. Walden, "Interactions in distance education: A report from the other side," from *1997 AERC Proceedings:* http://www.edst.educ.ubc.ca/aerc/1997/97burnham.html, 1997.

[5] J. M. Carman, "Blended Learning Design: Five Key Ingredients," from *Aglilant Learning*: http://www.agilantlearning.com/pdf/Blended Learning Design.pdf, 2005

[6] Centre for Learning and Professional Development. The University of Adelaide, Content Online, from *Online Education Service*: http://www.adelaide.edu.au/clpd/online/building/, 2008

[7] C. D. Dziuban, J. L. Hartman and P. D. Moskal, "Blended Learning," *Educause Center for Applied Research*, Research Bulletin , 2004 (7).

[8] R. Mason, "Models of Online Courses." in L. Banks, C. Graebner, & D. McConnel (Red.), *Networked Lifelong Learning: Innovative Approaches to Education and Training Through the Internet*, Sheffield: University of Sheffield, 1998.

[9] S. Mishra, *Designing Online Learning*, West Broadway: The Commonwealth of Learning, 2001.

[10] M. G. Moore, "Three Types of Interaction," *American Journal of Distance Education*, vol. 3 (2), 1989, pp. 1-16.

[11] B. Niemierko, *Pomiar wyników kształcenia*, Warszawa: WsiP, 1999.

[12] M. F. Paulsen, "Teaching methods and techniques for computer-mediated communication," *The 18th ICDE World Conference: The new Learning Environment, A Global Perspective*, ICDE, Pennsylvania State University, 1997.

[13] M. F. Paulsen, *Online Education and Learning Management Systems. Global E-learning in a Scandinavian Perspective,* Bekkestua: NKI Forlaget, 2003.

[14] L. Ragan, *Best Practices in Online Teaching,* Houston: Connexions, Rice University, 2008.

[15] A. Rossett, F. Douglis and R. V. Frazee, "Strategies for Building Blended Learning," from *Learning Circuits*: http://www.learningcircuits.org/2003/jul2003/rossett.htm, 2003

[16] C. H. Tu, "Critical examination of factors affecting interaction on CMC," *Journal of Network and Computer Applications*, vol. 23 (1), pp. 39-58, 2000.

[17] C. A. Twigg, "Improving Learning and Reducing Costs: New Models for Online Learning," *Educause Review Magazine*, vol. 38 (5), 2003.

[18] P. Valiathan, "Blended Learning Models," from *Learning Circuits*: http://www.learningcircuits.org/2002/aug2002/valiathan.html, 2002.

One Dimensional IIR Digital Filter Modeling Based on Recurrent Neural Network

Stefanova S.A.

Department of Electronic Engineering and Technologies, Technical University of Sofia,
8 Kliment Ohridski Str.,
Sofia, 1797, Bulgaria

***Abstract* - One approach for modeling of 1-D IIR digital filter based on two layer recurrent neural network is proposed. The Lagrange multipliers method has been applied to the training process of the neural network. The set of time domain data is generated and used as a target function in the training procedure. To demonstrate the effectiveness of the proposed neural network model some simulations have been done using input harmonic signals with different frequencies. The analysis of the behaviour of neural network model and target filter frequency responses shows good approximation results.**

I. Introduction

Neural networks have increasingly been used in many areas of signal processing. In last year various approaches for modeling and simulation of the analog and discrete dynamic systems based on different type of neural networks have been developed [1-9]. Basic results related to the discrete dynamical systems approximation using neural networks are discussed in [1]. Some of these methods are successfully applied to the design problem in time domain for analog [2] the non-recursive, recursive and adaptive digital filters [3-5]. One approach for the 1-D FIR digital filter design based on the weighted mean square method and neural network to state the approximation problem is proposed in [3]. Some methods for the non-linear digital filters design using neural networks are considered in [4]. One approach for time domain design of 1-D and 2-D recursive digital filter, based on recurrent neural network is proposed in [5].

A various methods have been developed for the purposes of the training process of the recurrent neural networks [6-9]. The backpropagation through time method is the basic approach for recurrent neural network training, discussed in [6]. Some training algorithms based on the sensitivity theory are considered in [7, 8]. Reference [5] describes the learning algorithm for the neural network structure based on the state space representation of 1-D and 2-D IIR digital filters.

In this paper an IIR or recursive digital filter model based on two layer recurrent neural network is proposed. The structure and mathematical description of the recurrent neural network is given in section II. The training process of the neural network using Lagrange multipliers method is discussed in Section III. A set of time domain data has been generated and used as a target function in the training procedure. The neural network model has been trained in such a way that with given predetermined input signal, the output variable approximates the target function in mean square sense. The neural network model of the recursive digital filter is considered in section IV. The effectiveness of the proposed neural network model of the recursive digital filter is demonstrated in Section V. The modeling results have been obtained for two different cases – a Nyquist recursive digital filter witch has a special type of the impulse response and a 4-th order bandpass recursive digital filter. Some simulations are realized using harmonic signals with different frequencies in the filter's passband and stopband. The analysis of the behavior of the neural network model and the target filter frequency responses shows good approximation results. The proposed approach can be extended and successfully applied to the case of modeling of nonlinear IIR digital filters.

II. Structure and Description of the Neural Network

The two layer recurrent neural network is considered. The structure of neural network is shown in Fig. 1.

The recurrent neural network is described with the recurrent system of equations as follows:

$$\begin{bmatrix} s_1(k) \\ s_2(k) \\ \cdots\cdots \\ s_n(k) \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} \ldots & w_{1n} \vdots \ldots w_{1n_z} \\ w_{21} & w_{22} \ldots & w_{2n} \vdots \ldots w_{2n_z} \\ \cdots\cdots\cdots & \cdots\cdots\cdots & \cdots\cdots \vdots \cdots\cdots \\ w_{n1} & w_{n2} \ldots & w_{nn} \vdots \ldots w_{nn_z} \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ \cdots\cdots \\ x_n(k) \\ u_1(k) \\ \cdots\cdots \\ u_{n_u}(k) \end{bmatrix},$$

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ \cdots\cdots \\ x_n(k+1) \end{bmatrix} = \begin{bmatrix} f(s_1(k)) \\ f(s_2(k)) \\ \cdots\cdots \\ f(s_n(k)) \end{bmatrix} \tag{1}$$

where f(.) is the activation function for the first layer of the neural network

Let be introduced the vector function $\mathbf{f}(\mathbf{s})$ as follows:

$\mathbf{f}(\mathbf{s}) = [f(s_1), f(s_2), \ldots, f(s_n)]^T$,

then the neural network description in the matrix form is:

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{W}\mathbf{z}(k)), \quad \mathbf{x}(k_0) = \mathbf{x}_0 \quad , \tag{2}$$

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) \tag{3}$$

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_52,

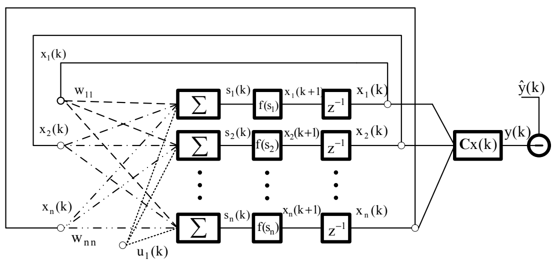


Fig. 1. Recurrent neural network structure for digital filter modeling

where

$\mathbf{x} \in \mathbf{R}^n$ is the state space vector,

$\mathbf{u} \in \mathbf{R}^{n_u}$ is the vector of neural network input excitations

$\mathbf{y} \in \mathbf{R}^m$ is the vector of neural network output,

$\mathbf{W} \in \mathbf{R}^{n \times n_z}$ and $\mathbf{C} \in \mathbf{R}^{m \times n}$, $m \le n$, $\mathbf{C} = \lfloor c_{ij} \rfloor$,

$c_{ij} = \begin{cases} c_i & i = j \\ 0 & i \ne j \end{cases}$ are the matrixes of system coefficients,

n - is the number of neurons at the first layer;

n_u - is the number of input excitations;

$n_z = n + n_u$ - is the number of neural network inputs.

The equations (2), (3) can be used to describe the recurrent neural network shown in Fig. 1.

The following additional vectors have been defined:

- a vector of the first layer inputs of neural network

$$\mathbf{z} = [\, x_1, x_2, \dots x_n, u_1, u_2, \dots, u_{n_u}\,]^T, \ \mathbf{z} \in \mathbf{R}^{n_z} \tag{4}$$

- a vector of the neural network outputs

$$\mathbf{y}(k) = [\, y_1(k), y_2(k) \dots y_m\,(k)]^T \tag{5}$$

- a vector of the neural network weighting coefficients

$$\mathbf{p} = \lfloor w_{11}\ w_{12} \dots w_{1n_z}, w_{21}\ w_{22} \dots w_{2n_z}, \dots w_{nn_z}, c_1, c_2 \dots c_m \rfloor,$$

$$\mathbf{p} \in \mathbf{R}^{n_p} \tag{6}$$

where $n_p = n.n_z + m$ is the number of neural network coefficients and the elements of matrix $\mathbf{W} \in \mathbf{R}^{n \times n_Z}$ are introduced row by row.

Mean square error objective function is defined in following form:

$$\mathrm{J}(\mathbf{p}) = \frac{1}{2} \sum_{k=k_0}^{k_f - 1} \sum_{i=1}^{m} [\mathrm{y_i}(k) - \hat{\mathrm{y}}_i(k)]^2 \tag{7}$$

where $\{\hat{y}_i(k)\}$ is a set of samples of the target function.

III. Algorithm of Lagrange Multipliers

The main problem in the neural network training process is the gradient calculation of the mean square objective function (7) with respect to weights. The algorithm of Lagrange multipliers [9] is used as a training procedure of the neural network model. The Lagrange multipliers algorithm has been written in matrix form.

The vectors of Lagrange multipliers are defined as:

$\boldsymbol{\lambda}(k) = [\lambda_1(k), \lambda_2(k), \dots, \lambda_n(k)]^T$;

$\boldsymbol{\Gamma}(k) = [\Gamma_1(k), \Gamma_2(k), \dots, \Gamma_{n_p}(k)]^T$;

and the Hamiltonian of the optimization problem (2),(3), (7) is stated in the form:

$$H = \frac{1}{2} \sum_{i=1}^{m} (y_i(k) - \hat{y}_i(k))^2 + \boldsymbol{\lambda}^T(k+1)[\, f(\sum_{j=1}^{n_z} w_{1j} z_j), \\ f(\sum_{j=1}^{n_z} w_{2j} z_j), \dots, f(\sum_{j=1}^{n_z} w_{nj} z_j)\,]^T + \boldsymbol{\Gamma}^T(k+1)\mathbf{p} \tag{8}$$

Using (8) the conjugated system is composed as follows:

$$\boldsymbol{\lambda}(k) = \frac{\partial H}{\partial x(k)} \qquad \boldsymbol{\lambda}(\kappa_f) = 0; \tag{9}$$

$$\boldsymbol{\Gamma}(k) = \frac{\partial H}{\partial \mathbf{p}(k)} \qquad \boldsymbol{\Gamma}(\kappa_f) = 0. \tag{10}$$

The gradient of the objective function (7) can be calculated using the following algorithm:

Step 1. Calculate and store the set of values $\{\mathbf{x}(k)\}$ from (2) for $\mathbf{x}(k_0) = \mathbf{x}_0;\ k = k_0, k_0 + 1, \dots, k_f - 1$.

Step 2. Solve the conjugate system (9) и (10) for $k = k_f - 1, \dots, k_0$ (backwards in time).

Step 3. Obtain the objective function gradient from the solution of (10) for $k = k_0$

$$\nabla J(\mathbf{p}) = \boldsymbol{\Gamma}(k_0).$$

The conjugated system (9), (10) can be written in matrix form. Then it is necessary to define the following matrix and vectors:

- a sub-matrix $\mathbf{W}_x$ of the weighting coefficients matrix $\mathbf{W}$:

$$\mathbf{W}_x = \begin{bmatrix} w_{11} & w_{12} \cdots & w_{1n} \\ w_{21} & w_{22} \cdots & w_{2n} \\ \cdots & \cdots & \cdots \\ w_{n,1} & w_{n,2} \cdots & w_{n,n} \end{bmatrix}; \tag{11}$$

- a matrix of Lagrange multipliers for the first layer of the neural network:

$$\boldsymbol{\Gamma}^w(k) = \begin{bmatrix} \Gamma_{11}(k)\,\Gamma_{12}(k) \dots \Gamma_{1,n_z}(k) \\ \Gamma_{21}(k)\,\Gamma_{22}(k) \dots \Gamma_{2,n_z}(k) \\ \cdots \\ \Gamma_{n1}(k)\,\Gamma_{n2}(k) \dots \Gamma_{n,n_z}(k) \end{bmatrix}; \tag{12}$$

- a vector of Lagrange multipliers for the second layer of the neural network:

$$\boldsymbol{\Gamma}_c(k) = [\Gamma_{c1}(k)\,\Gamma_{c2}\,, \dots, \Gamma_{cm}]^T; \tag{13}$$

- a diagonal matrixes

$$\mathbf{D}_1 = diag(\, f'(s_1(k))\, f'(s_2(k)), \dots, f'(s_n(k))\,)$$

$$\mathbf{D}_2 = diag(\, x_1(k)\, x_2(k), \dots, x_m(k)\,). \tag{14}$$

The error vector from the objective function (7) can be written as:

$$\mathbf{e}(k) = [e_1(k), e_2(k), ..., e_m(k)],$$
$$e_i(k) = y_i(k) - \hat{y}_i(k), \quad i = 1,2,...,m \tag{15}$$

Equation (9) can be stated in matrix form as follows:

$$\boldsymbol{\lambda}(k) = \mathbf{C}^T \mathbf{e}(k) + \mathbf{W}_x^T(k)\mathbf{D}_1\boldsymbol{\lambda}(k+1),$$
$$\boldsymbol{\lambda}(k_f) = 0, k = k_f - 1,...,k_0 \tag{16}$$

where $\mathbf{C}$ is the matrix from (3), $\mathbf{W}_x$ is the matrix from (11), $\mathbf{D}_1$ is the matrix from (14) and $\mathbf{e}(k)$ is an error vector (15).

Equation (10) can be stated in the following form:

- for the first layer of the neural network:

$$\boldsymbol{\Gamma}^w(k) = \boldsymbol{\Gamma}^w(k+1) + \mathbf{D}_1\boldsymbol{\lambda}(k+1)\mathbf{z}^T(k), \quad \boldsymbol{\Gamma}^w(k_f) = 0,$$
$$k = k_f - 1,...,k_0 \tag{17}$$

where $\boldsymbol{\Gamma}^w(k)$ is the matrix from (12);

- for the second layer of the neural network:

$$\boldsymbol{\Gamma}_c(k) = \boldsymbol{\Gamma}_c(k+1) + \mathbf{D}_2\mathbf{e}(k), \quad \boldsymbol{\Gamma}_c(k_f) = 0 \tag{18}$$

where $\boldsymbol{\Gamma}_c(k)$ is the vector from (13);
$\mathbf{D}_2$ is the matrix from (14).

The gradient of the objective function (7) for the conjugated system (16), (17), (18) has been calculated applying the three steps algorithm, described above, as follows:

$$grad\ J(\mathbf{p}) = \left[\boldsymbol{\Gamma}_1^w(k_0)\,\boldsymbol{\Gamma}_2^w(k_0)...\boldsymbol{\Gamma}_n^w(k_0)\,\boldsymbol{\Gamma}_c^T(k_0)\right]^T,$$

where $\boldsymbol{\Gamma}_i^w(k)$, $i-th$ row of the matrix $\boldsymbol{\Gamma}^w(k)$, *i =1,2,..., n*.

The statement of conjugated system (16), (17) in a matrix form allows effective realization of the algorithm using to determine the gradient of the objective function (7) with respect to neural network weights. Standard gradient optimization procedure has been applied to minimize the objective function (7) that is defined for the solution of the system (2), (3).

IV. NEURAL NETWORK MODEL OF IIR DIGITAL FILTER

The IIR or recursive digital filter can be considered as a linear discrete dynamic system described in time domain with n order difference equation, which has been stated using the delayed samples of the input excitation and the response at the output. Transforming the difference equation the state space description of the recursive digital filter can be obtained as a linear system of n first order recurrent equations. The advantages of the state space representation are the simplicity of time domain analysis, the matrix form description of digital filter impulse and step responses and possibility of parallel calculations.

The neural network structure shown in Fig. 1 must be modified to linear structure. In this case the activation function f(.) is linear. This linear network is used to create a model of recursive digital filter. The corresponding system of equations (1)-(3) is a linear recurrent system. The Lagrange multipliers algorithm modified for linear network is used as a training procedure of the neural network model. The training sequences are generated using the impulse response of target recursive digital filter that is designed by Filter Design Toolbox of MATLAB. The corresponding excitation at the input of the neural network model is unit impulse.

V. MODELING RESULTS

Case 1. Modeling of 4 th order bandpass IIR digital filter

The following requirements to the magnitude response of the target bandpass digital filter are specified: bandpass [300–3400] Hz, passband loss – 3 dB; bandstops [0–100] Hz and [3600–4000] Hz, bandstop loss – 15 dB; sampling frequency – 8000 Hz.

The model of the bandpass IIR filter is realized as recurrent neural network structure with 4 neurons.

The effectiveness of the proposed algorithm is demonstrated by simulation with neural network model. The simulation is realized applying input harmonic excitations with frequencies in the digital filter's passband and stopband. These sinusoidal signals are different from the set of data using in the training process of neural network model. The impulse responses of the target digital filter and the neural network model are shown in Fig. 2 and Fig. 3 and the magnitude responses in frequency domain are shown in Fig. 4 and Fig. 5, respectively.

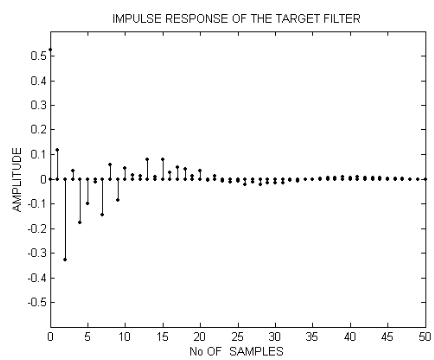


Fig 2. Impulse response of the target bandpass IIR filter

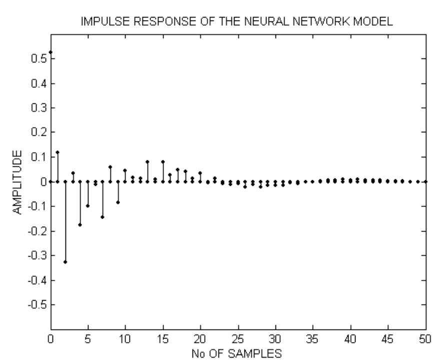


Fig. 3. Impulse response of the neural network model

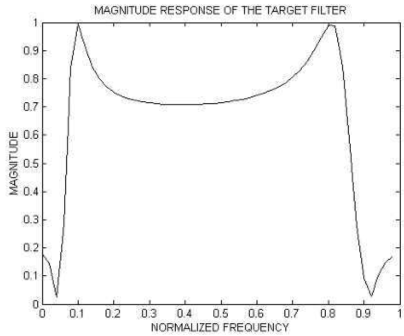


Fig 4. Magnitude response of the bandpass target filter

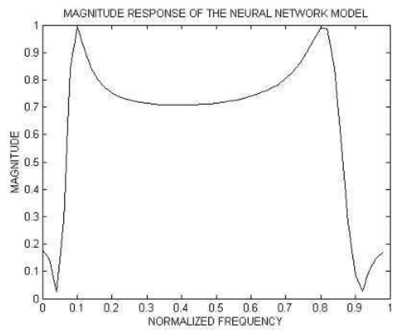


Fig. 5. Magnitude response of the neural network model

The simulation experiments are implemented when at the input of the neural network model have been applied sinusoidal signals with the following frequencies - 50 Hz from the filter's stopband and 1000 Hz – from the filter's passband. The comparison between the input harmonic excitation and the output signal of neural network model is illustrated in Fig. 6 – the first case of a 50 Hz input harmonic signal and Fig. 7 – the second case of a 1 kHz input harmonic signal. The time responses of the target bandpass digital filter and the neural network model in the case of input signal - 50 Hz and 1000 Hz are respectively demonstrated in Fig. 8 and Fig. 9.

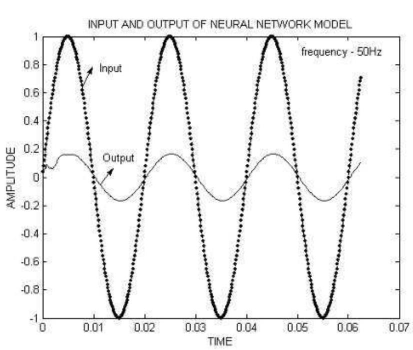


Fig 6. Input and output of the neural network model, case 1- 50 Hz

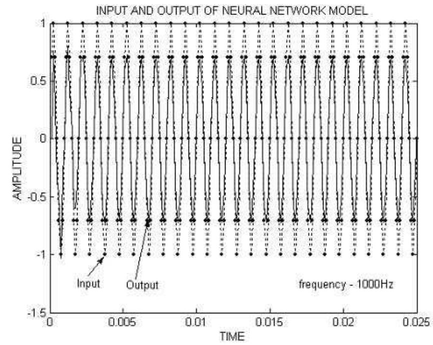


Fig 7. Input and output of the neural network model, case 2 - 1 kHz

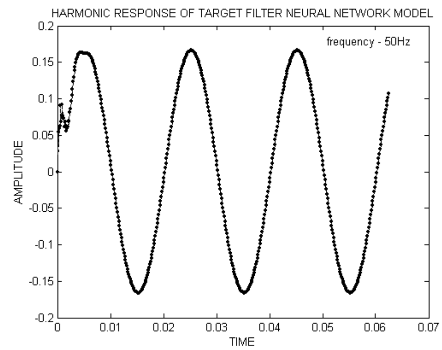


Fig. 8. Harmonic responses of target filter and neural network model – 50 Hz

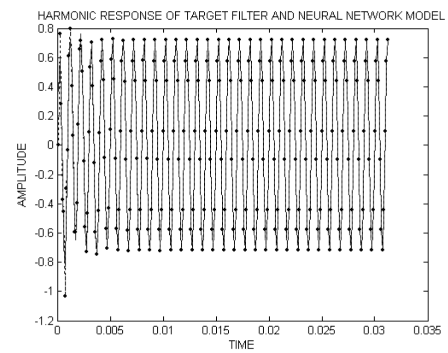


Fig. 9. Harmonic responses of target filter and neural network model – 1 kHz

Case 2. Modeling of Nyquist recursive digital filter

The Nyquist recursive digital filter plays an important role in digital data transmission for its intersymbol interference (ISI)-free property. Also it can be adopted in decimation or interpolation multirate systems. To achieve zero ISI, Nyquist filter must satisfies some criteria in time domain that they should have zeros equally spaced in the impulse response coefficients except one specified. Infinite impulse response (IIR) Nyquist filter has lower orders than FIR filters, but their impulse responses are more difficult to keep the zero-crossing time constraint property ant the problem of filter stability should also be examined [10].

Impulse response $h(n)$ of the (IIR) Nyquist filter with the time domain constraints is defined in the form [10]:

$$h(K+kN)=\begin{cases}\dfrac{1}{N}\neq 0\,, & if\ k=0\\ 0\,, & otherwise\end{cases} \tag{19}$$

where K and N are integers

The transfer function $H(z)$ of the (IIR) Nyquist filter can be expressed as:

$$H(z)=b_K z^{-K}+\frac{\sum\limits_{i=0,\,i\neq kN=K}^{N_n} b_i z^{-i}}{\sum\limits_{i=0}^{N_d/N} a_{iN} z^{-iN}}\,,\quad b_K=\frac{1}{N} \tag{20}$$

where N_n, N_d are integers, all the filter coefficients a_i, b_i are real, $a_0=1$, N_d is the multiple of N.

The impulse response of the IIR Nyquist filter is used as target function in the training procedure of the neural network model. The IIR Nyquist filter with $N_n=15, N_d=4$, $N=4, K=9$ is considered.

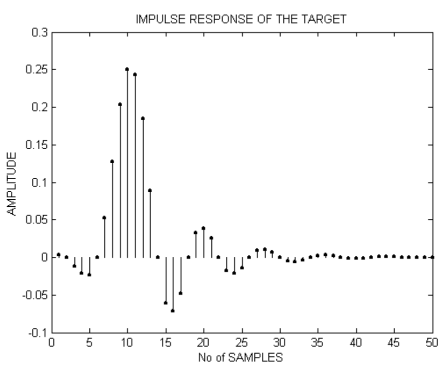


Fig 10. Impulse responses of the target Nyquist IIR filter

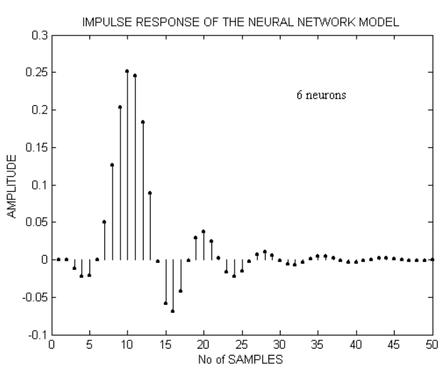


Fig. 11. Impulse response of the neural network model

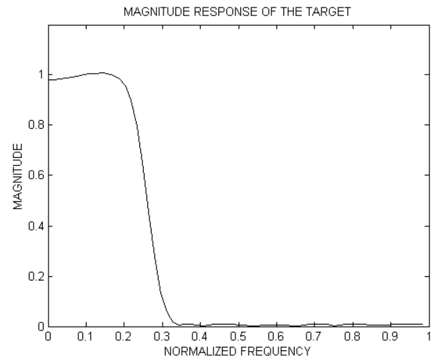


Fig 12. Magnitude responses of the target IIR filter

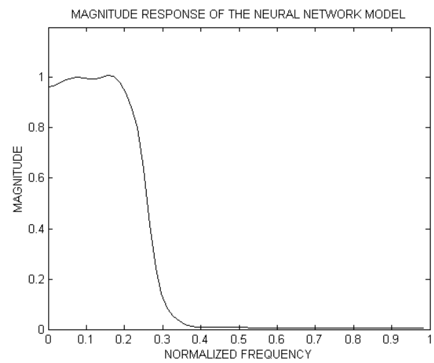


Fig. 13. Magnitude response of the neural network model

The impulse responses of the target IIR Nyquist filter and the neural network model with 6 neurons are shown in Fig. 10 and Fig.11. The magnitude responses in frequency domain of the Nyquist filter and the neural network model with 6 neurons are shown in Fig. 12, Fig. 13.

VI. CONCLUSION

One approach for the IIR digital filter modeling based on recurrent neural network is proposed. The Lagrange multipliers method has been used to create an effective training algorithm for the neural network model. The Nyquist and bandpass IIR digital filters have been modeling using the recurrent neural network structures. To investigate the behavior of the digital filter models some simulations have been realized using input harmonic excitations with different frequencies. The analysis of the obtained results shows a very good approximation of the impulse responses in time domain and the magnitude responses in the frequency domain. This approach can be extended and successfully applied to the case of modeling of nonlinear recursive digital filters.

REFERENCES

[1] Jin L., P. Nikiforuk, and M. Gupta, "Approximation of Discrete-Time State-Space Trajectories Using Dynamical Recurrent Neural Networks", *IEEE Transactions on Automatic Control,* Vol.40, No 7, 1995.

[2] H. Cruse , *Neural Networks as Cybernetic Systems*, Brains, Minds and Media, Bielefeld, Germany, October 2006.

[3] D. Bhattacharya, and A. Antoniou, *Real-time Design of FIR Filter by Feedback Neural Network*, IEEE Signal Processing Letters, Vol.3, No 4, May 1996, pp. 158 – 161.

[4] M. Pedersen, "Optimization of Recurrent Neural Networks for Time Domain Series Modeling", *Ph.D Thesis, Department of Mathematical Modeling*, Technical University of Denmark, 1997.

[5] Wang D. and A. Zilouchian, *Intelligent Control Systems Using Soft Computing Methodologies*, Editors: Ali Zilouchian, Mo Jamshidi**,** Prentice Hall, March 2001, pp, 93-110.

[6] W. T. Miller, R. S. Sutton and P. J. Werbos, *Neural networks for Control*., Cambridge, MA:MIT Press, 1990.

[7] R. Grino, G. Cembrano and C. Torras, "Nonlinear system identification using additive dynamic neural networks – two on-line approaches", *IEEE Trans. Circuits and Systems – I*, vol. 47, no. 2, 2000, pp. 150-165.

[8] Y. Fang, M. Yagoub, F. Wang and Qi-J. Zhang, "A new macromodeling approach for nonlinear microwave circuits based on recurrent neural networks", *IEEE Trans. on Microwave Theory and Techniques*, Vol. 48, no. 2, Dec. 2000, pp. 2335-2344

[9] A. P. Sage and J. L. Melsa, "System Identification", vol. 80 of *Mathematics in Science and Engineering*, New York, NY:Academic, 1971.

[10] S. Pei, Ch. Hsu, P. Wang, "Design of a class of IIR eigenfilters with time and frequency domain constrain*t, IEEE Trans. on Circuits and Systems-II Analog and Digital Signal Processing*, vol.49, No 2, February 2002, pp 145 - 151.

The Development of a Lecture Capture System Based on a Tool to Support Hearing Impaired Students

Iain Stewart
Glasgow Caledonian University
Cowcaddens Road
Glasgow
Scotland

Malcolm Allan
Glasgow Caledonian University
Cowcaddens Road
Glasgow
Scotland

David Harrison
Glasgow Caledonian University
Cowcaddens Road
Glasgow
Scotland

***Abstract*- In this paper, we describe the development of a support tool based on voice recognition technology for deaf students in the lecture environment. The successful prototype is then extended in a pilot scheme to provide additional material for the entire student body. Based on the feedback from the pilot scheme a proposal for a framework to support student interaction with captured lecture content is developed. The work is placed in the context of other developments and issues in the fields of deaf studies and e-learning.**

I. Introduction

Across the educational sector there has been an increased drive to increase the accessibility of education. In the UK where this project is based the key pieces of legislation are the Disability Discrimination Act [1] and the Special Educational Needs and Disability Act [2].

As part of the lecturing work by the authors, a significant number of deaf students were encountered in the classroom environment. Traditional support mechanisms were applied to ensure that they received a good quality education but the authors realized that the multimedia technologies being taught could be used to enhance the learning experience for these deaf students. The national figure for the number of deaf students entering higher education is approximately 2.6 for every 1000 [3]. This figure was significantly lower than that encountered by the authors who had 4 deaf students out of a programme cohort of 210 in a 5 year period.

The particular problems that the students encountered were in dealing with subjects with high technical jargon content as this is the area where the traditional support tools are least effective.

II. Support Issues for Deaf Students in Lectures

The problems faced by deaf students in higher education have been the subject of considerable investigation. Researchers such as Fleming & Hay [4] and Bremner & Housden [5] have identified a range of issues that deaf students face in this area. One area that is identified as requiring further investigation is that of the role and effectiveness of the support services in improving academic performance [6]. From these and other works, it has been established that the communications mechanism between the lecturer and the deaf student is one of the key areas that affect performance. In particular the lecture theatre environment is identified as an area where the information exchange is not particularly effective.

There are a range of factors which impact on the ability of the deaf student to gain the maximum benefit from the lecture process. In the lecture context, the most common and effective form of support is the sign language interpreter (or signer). This person is highly skilled and there is therefore a great demand for their services. In Scotland for example it is estimated that over 300 full time signers would be required to meet the need of the community. The figure for 2003 was that there were 42 registered interpreters. [7]

The second major issue when working with signers is that the signers very rarely have the subject specific knowledge and sign vocabulary to translate the technical terms used in the lecture into sign language. If this specialist vocabulary cannot be developed then the signer has to resort to translating via finger spelling which is a slow process and quickly falls out of step with the lecture content. Recently an online reference resource called "Sciencesigns", has been developed which provides a British Sign Language (BSL) glossary for Engineering and Science [8], [9].

The importance of sign language translation from specialist signers is supported by a range of researchers who have shown that "…deaf students readily acknowledge that they do not receive full access to information in university lectures." [10]. In particular it has been shown [11] that deaf students felt the need for subject specific knowledge in their interpreters to improve their comprehension of the topic. This was also supported by the work of Quinsland and Long [12] who reported that the immediate retention of lecture content was twice as high if the lecture was translated by a specialist signer.

It is possible for the lecturers themselves to become trained in signing but it is the experience of one of the authors (having completed BSL level 1) that the level of proficiency required to fully support teaching would form a significant additional load for most staff. In addition, most interpreters use sign language as a second language "As a result, the spoken language tends to dominate most cognitive functioning." [13]. This results in a poor presentation of the concepts in BSL together with the

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_53,

cognitive load for the lecturer of supporting the concurrent speech and signing activity.

Marschark et al [14] reviewed a range of surveys into deaf education and identified that there were a number of factors which limited the ability of deaf students to benefit fully from the classroom experience. One key finding was that the overall level of comprehension of interpreted content was 85% of that of the hearing students. An interesting additional point from the work was that the authors identified that there were an increasing number of deaf students who were requesting access to online text presentations as an alternative to interpretation.

This outcome, together with the ongoing shortage of technically specialized sign language interpreters suggested that an alternative support tool would be worth developing.

The approach that was decided upon was to apply speech recognition technology to capture the lecture "live" and to provide this text in near real time as subtitles on the presentation screen.

III. THE DEVELOPMENT OF THE TALKSHOW PROJECT

When Talkshow project was initiated (2001), speech recognition technology had advanced sufficiently that it was now possible to run the software on a high end desktop system.

One of the core principles of the Talkshow project was to try and produce a low cost solution to support deaf students in the lecture and group working environments. It was intended that it would be possible to deploy the system with the type of resources that would normally be available to the general teaching staff in a university and without dedicated technical support. The project integrated an off the shelf voice recognition package with an application which displayed the transcribed text in subtitle mode under the presentation. [15]. (Fig. 1)

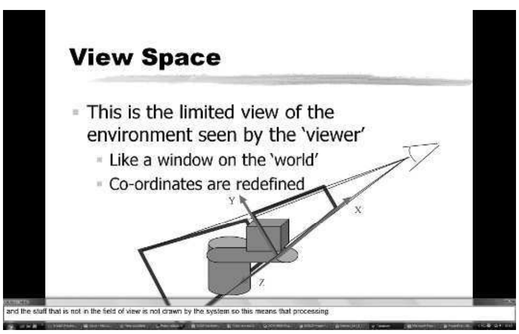


Fig 1. A screen capture of Talkshow in use.

Other products which took advantage of the advances in speech recognition to support deaf students were being developed at the same time as the initial Talkshow project. The most notable example of these was the Liberated Learning Consortium [16], but at that time the cost, complexity and practicality of the solutions reduced their impact and deferred implementation [17].

A major assumption which underlies the Talkshow project was that the deaf students who would use it were fluent in written English. It cannot always be assumed for profoundly deaf students that their written English skills are of a high level. This is because most deaf people who have been deaf from birth learn English as a second language, after BSL.

The critical role that written English plays in the higher education of deaf students and in particular to be able to access many of the technical resources and assistive technologies is is highlighted by a quote from Jordan [18] which states "…in order to take advantage of the emerging technologies, deaf people have to be highly literate. They must be able to read and write fluently! It is essential that students become highly skilled in reading and writing English….. So obviously literacy is a high priority".

At present the experience of the authors is that those deaf students who study in the technological fields do have such levels of English, so the issue of lack of written English skills did not impact on the research activity.

This situation is changing both socially and technically. A significant example of this in the UK was the recognition in 2003 that British Sign Language was a language in its own right within the United Kingdom [19]. In the technical context, work is underway on the development of video based translation [20] and signing avatars which can convert written text into sign language [21].

At present these developments are in their early stages and can usually only translate into sign supported English as BSL has a different grammar to English.

IV. RESULTS FROM THE TALKSHOW PROJECT

By 2005 the Talkshow prototype evolved to provide support for group working as well as lectures but for this paper the focus will remain on the lecture support element of the project.

The results obtained from the project showed that while the project had definite limitations, in general the deaf students found it to be a useful addition to the support that was available to them [22]. Key points were:

- The speech recognition was sufficiently accurate to provide useful additional information to the deaf students.
- In most cases the content of the subtitles was found to be accurate and allowed the student to follow the lecturer's speech effectively.
- The deaf students were already used to having to read subtitles and multi channel information.
- When the text produced was garbled, it was easily identified by the students and ignored.

- Technical terms and digressions which were used to illustrate the meaning of these terms were easier to pick up
- Questions from the class (if echoed by the lecturer) were also able to be followed, increasing the student's understanding of the topic and making the student feel more integrated into the lecture class.

One particular benefit was that the system was most effective in the area of technical terminology where the use of signers had the most problems. Assuming the lecturer had taken time to train the system in the technical vocabulary of the lecture, the technical terms which were identified as being the most difficult for the signer to translate effectively were normally translated successfully by the speech recognition engine. It was identified that the unusual nature of the terms meant that there was less likelihood of a translation mismatch. Researchers at MIT [23] did not find this to be the case but the difference in the results was identified as being down to the time that the lecturer spent prior to the lecture in ensuring that the vocabulary was recognised by the system.

The key area of translation accuracy is one of the areas that limits the usefulness of voice recognition. The performance of the system is dependent on a range of factors including:

- The quality and positioning of the microphone.
- The acoustics of the room being used.
- The amount of time taken to train the voice model
- The changes in the speech patterns and voice characteristics in a lecture environment
- The processing power of the system.
- The existence of the words in the vocabulary of the system (many technical terms have to be added during training.)

Speech recognition works with phrases to work out the context and hence the words to select from the vocabulary. If a word is mistranslated then this can result in the system losing the sense of the phrase or sentence.

This is supported by researchers such as Glass et al [24] who found when using a standard, unidirectional microphone, attached to a video camera, to capture the content of the lecture, that transcription errors of the order of 30% to 40% were to be expected. This would be unacceptable for use in lectures and would require significant error correction at a later date, with the related time penalty. When using a dedicated noise cancelling headset microphone, other researchers in the area, notably Wald [25] found that error rates of 10% to 15% were the norm for the lecture environment. This tied in with the experience of the authors.

As part of the evaluation of the project, hearing students in the class were also asked to evaluate the usefulness of the prototype. The general feedback from the hearing students was that while they found the application to be an interesting novelty in the lecture, the subtitling was a distraction for several reasons.

- There is an appreciable delay between the speech and the phrase being translated which resulted in a discontinuity between the written and spoken information;
- The continuously scrolling text became a visual distraction;
- When the recognition engine did mistranslate the content of the speech the results were often humorous and this disrupted the flow of information.

As a result of this feedback the later utilization of Talkshow was set up so that a local monitor was made available to the deaf students for them to refer to while the main lecture was projected without the subtitles onto the screen

Many of the hearing students did request that the transcript of the lecture generated by Talkshow be made available to them to support their revision work.

The lecturing staff who worked with the prototype identified a range of issues with the technology. In particular:

- The training time required for effective translation was considered to be too great in some cases.
- All staff who used the prototype felt that it was cumbersome. This has now been improved by the use of wireless headsets.
- Opinion was split over whether the technology had a positive or negative impact on the lecturer's delivery.

While Talkshow was not a replacement for a skilled signer, well practiced in the technical terms of the subject matter, when this ideal was not available, the effectiveness of communication of the lecture content was improved for deaf students. The product still had limitations in terms of ease of use and the training required to achieve proficiency in its use.

V. Extending the Tool to Form a Prototype Lecture Capture Resource

The feedback from the hearing students from the Talkshow pilot project suggested that there was an opportunity to extend the research activity. It was decided to investigate whether the transcribed lecture content would provide a useful additional resource which could be integrated into the learning environment of the wider student body. In particular the intention was to integrate the transcribed lecture content with audio, video and screen captures of the presentation to provide a rich resource for the students to use to support their learning.

It is now common for lectures and lecture content to be made available to the students in a variety of formats. The normal minimum level of provision is via Managed Learning Environments (MLEs) are used by the lecturer to provide flexible access to lecture content, typically presentation slides and notes [26]. This is extended by a range of commercial and research projects which provide the capability to capture more of the lecture. Typical examples of these are ECHO360 [27] and

Mediasite [28]. The general characteristic of these technologies is that they automate the live capture of lectures and make it easy to place the resultant content online. This allows students to review the lecture in their own time or from a distance. The challenge with this technology is to make the captured content easy to navigate and to manipulate. One significant project in this area is the collaboration between MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) [23] and Microsoft to develop techniques for the transcribing and indexing of academic lectures.

Major strides have been made in the indexing and navigation of captured lectures but identifying and moving quickly to a particular section can still be dependent on navigating in a manner similar to DVD chapters, [29] assuming that the captured presentation has slides and video/audio content, and it is difficult to search the lecture for a particular element, whereas text is relatively easy to search.

A significant additional benefit of incorporating the captured text is that, unlike video or audio, it is easy to amend and correct errors or incomplete elements of the lecture presentation.

In terms of making the lecture content more accessible to the general student body and in particular to students with disabilities, the captured text helps to reduce the cognitive burden of notetaking [30]. By removing this burden, or at least reducing it to the annotation of particular key points in the lecture the ability of the student to concentrate on the content of the lecture is increased. If the content is made available or produced during the "live" presentation then it is possible for the student to add annotation to the lecture content in real time, particularly with the wide availability of low cost wireless laptops within the student body. Deaf students in particular are often provided with grants to purchase such devices. The potential of this area has been recognised by researchers and developments to take advantage of the live content have been proposed [31].

This work also ties in with the work of other developers of educational software and e-learning technologies. These systems include sophisticated lecture capture systems such as that under development at Microsoft [32].

It is generally recognised from these and other studies, that as it becomes possible to provide access to a wide range of captured media from lectures and presentations, some mechanism is desirable to provide the presenter with practicable means to develop content which will allow the student to structure and index the content to support their own learning style is desirable.

VI. Results from the Prototype

In order to investigate the effectiveness of this concept, a prototype application based on Talkshow was developed and deployed in a series of six honours level lectures in 2007. The lectures had a significant technical content. The purpose of this activity was to generate feedback from the students and to inform the development of the specification of the formal software project. To avoid any possible ethical issues, classes were chosen which did not have any deaf students. In this situation, Talkshow was simply used as the audio transcription tool. The captured text was then edited to remove errors and irrelevant content. The resultant text file was then mounted on the MLE along with the captured audio from the lecture and the PowerPoint slides.

The views of the students and the staff were collected and reviewed in order to inform the next stage of the development of the prototype. The students found that the ability to access the corrected text of the lecture was indeed a useful resource. By its nature, text is relatively easy to skim over or to search using software tools. It also made it possible for students to store the content on devices with limited memory and to print out the transcript for later study in conjunction with the slides.

The limits of the text as it stood were that it did still contain some errors (even after correction) and that as lectures are relatively free form spoken "performances", then there were times when the core content of the lecture could be difficult to follow. An initial problem with the text sometimes being difficult to relate to the presentation slides was at least partly addressed by incorporating it into the speakers notes section associated with each slide of the presentation file.

A particular view of some students was that the text was an especially useful additional resource, as it supported learning in locations where IT resources may not be available. It was possible to print out the slides and the related text and read over them while travelling in to university for example. The major criticism of the content was that it was another means by which they could be presented with the same information, there was no real opportunity to extend or expand the information and it was still a passive rather than active learning experience. From discussions with the students it became clear that many of them would find it even more useful to be able to interact with the content and thus use it more effectively in their studies.

From the staff perspective, the effort required to make the content useful (between 1 and 2 hours on average) was a significant overhead and they were not sure if the effort had sufficient educational benefit as it stood.

VII. The Specification of an Interactive Captured Lecture Framework

Based on the feedback from staff and students the following key points were identified as being essential to the extended lecture capture resource. In particular the specification had to recognise that the "one size fits all" approach to the capture and presentation of lecture materials does not allow for the wide variety of different learning styles and strategies that students use in their learning [33]. The particular challenges in the specification are firstly to develop a structured yet adaptable framework for the lecture content that will allow easy navigation of the content. At the core of this structure will be a timeline, captured by the presenter or inserted later, to which the various elements are linked. Secondly, as the transcribed text is an integral part of the specification, it is important that it be integrated into the system in such a manner as to allow easy modification, adaptation and linkage to the supplied media elements. The key features of the specification were:

- The system should be easy for staff to use and should not impact significantly on the ability of the lecturer to present the material in a manner that suits their lecture style.
- The presented content should be available in a variety of formats suitable for a range of delivery platforms.
- The content should be configurable by the student
- The content should be able to be added to either by staff or students. Examples of this kind of collaborative activity would include:
  - Clarification of topics and expanded definitions of terminology.
  - Links to supporting media provided at the appropriate point in the transcript.
  - Links to tutorials.
  - Questions raised by the students and the responses from the lecturer or fellow students.
  - Annotations and comments from individual users.

In order to allow this to be done the various items of captured lecture content should be presented in a common framework, with a common timeline to allow navigation through the material. By incorporating captured speech into a flexible e-learning environment the usability of the environment for both deaf and hearing students would be improved and would enhance the accessibility of the educational content.

Keyboard and other interactions would need to be captured and embedded in a timeline to allow linking of content. The final content should allow the individual staff member or student to modify and adapt the content to suit their own learning style and to increase the student's engagement with the content. The core of the specification is the development of a markup language (Talkshow Markup Language – TSML) that will allow the various elements of the lecture to be linked and manipulated via a common framework. This proposal ties in with the desire of students to work with web 2.0 technologies [34], which are an increasing element of their online environment. The authors of the paper sound an important note of caution as the increased user control over the structure and depth of their learning experience may result in some students choosing strategies that are ineffective or counter-productive.

At present the appropriate software platform to support the development and distribution of the content is being investigated and the design of the learning framework is under way, based on the running Talkshow prototype.

By providing a suitable framework then the lecture content and its related resources can be accessed in a variety of ways. For mobile learners a stripped down text and still image resource may be suitable whereas those working within a wireless campus environment may choose to access the content in a number of different modes, depending on their preferred learning style.

One of the key benefits of the structure under development is that it is intended to encourage greater student involvement with the content of the course. By allowing the students to take "ownership" of the content and to manipulate it in a manner that is appropriate to their learning style and mode of access, it is hoped that the engagement of the students with the content will be increased.

The framework is intended to improve the quality of the distance access provision as well. In particular, by incorporating an interactive element into the framework it allows questions and issues to be raised after the lecture and then linked back into the lecture content

VIII. Conclusion

In many cases lecturers already create a framework onto which a range of related materials can be mounted and adapted so as to provide a basis for a structured learning activity, some including captured lecture material. The fundamental change promoted by this proposal is that this needs to be more explicit and adaptable to encourage wider adoption of these resources and to convert the process of lecture capture away from simply recording the content of a lecture and towards creating an interactive, flexible learning resource to better support the student's learning and understanding. In order for this to be successful it is critical that the students continue to be involved with the development of this framework. To this end the next stage of this project is to concentrate on a particular honours level module with significant complex technical content. The content presented during the lectures and tutorial sessions will be captured as before but integrated into a more sophisticated framework structure than that currently used. This will allow the students and staff to evaluate the effectiveness of the revised framework and to further refine and extend the specification. Clearly, given the origins of the project, it is essential that the goal of improving the accessibility of the learning resource for all students must be core to the development of the project.

Acknowledgment

The authors would like to thank the colleagues and students (both deaf and hearing) who helped to test the various versions of the project for their support, tolerance and feedback.

REFERENCES

[1] The Disability Discrimination Act (c. 50) HMSO, 1995 ISBN 0 10 545095 2

[2] The Special Educational Needs and Disability Act, HMSO 2001 ISBN 0 10 541001 2

[3] M. Brennan, M. Grimes and E. Thoutenhoofd, “Deaf Students in Scottish Higher Education - A Report for the Scottish Funding Council 2005”, p11, Forest Books Ltd., 2007. ISBN – 0-94625-265-3

[4] J. Fleming, J. Hay, “Learning strategies of Deaf and hearing impaired students in higher education; CELT Learning and Teaching Projects 2002/03” 2003 ISBN: 0954211626

[5] (1996) A. Bremner, T. Housden, Issues in educational settings for deaf students and interpreters [online]. Retrieved March 24, 2002, from http://www.deakin.edu.au/extern/rdlu/deafstud.html.

[6] H. Lang, “Higher Education for Deaf Students: Research Priorities in the New Millennium” Journal of Deaf studies and Deaf Education, Vol 7 No 4 pp 267 – 280, Fall 2002 Online ISSN 1465-7325 - Print ISSN 1081-4159

[7] L. Lawson, “Scottish Council on Deafness”, IEE Symposium Digest on Assistive Technology- Towards a better life, pp 40-42 , 2003

[8] J. Mole, D. Peacock, Keynote speech. “Supporting Deaf Engineering Students – Issues and Opportunities for Learning and Teaching” Proceedings of the International Conference on Innovation, Good Practice and Research in Engineering Education. 2004 ISBN 0-9542116-5-0

[9] (2006) J. Mole. SignsOnline: British Sign Language (BSL)/English glossaries for Science and Engineering Education. [online] http://www.artsigns.ac.uk Accessed 22/1/2006

[10] J. Napier and R. Barker Accessing University Education: Perceptions, Preferences, and Expectations for Interpreting by Deaf Students, Jnl. of Deaf Studies and Deaf Education ,Volume 9, Number 2 pp. 228-238 Spring 2004; 9: 228 – 238 Online ISSN 1465-7325 - Print ISSN 1081-4159

[11] R. Locker, Lexical equivalence in transliterating for deaf students in the university classroom: Two perspectives. Issues in Applied Linguistics, Volume 1, number 2, pp 167–195. 1990 ISSN 1050-4273

[12] L.K. Quinsland, G. Long, Teaching, interpreting and learning: Implications for mainstream hearing impaired students Paper presented at the 1989 Convention of the American Educational Research Association, San Francisco, March. 1989

[13] CDS The Centre for Deaf Studies, at Bristol University, http://www.bris.ac.uk/deaf/library/deafstudies_info Accessed 4/2/02

[14] M. Marschark, P. Sapere, C Convertino & Seewagen R. Chapter 3 Educational Interpreting: Access and outcomes pp57 – 83 M. Marsharck, R. Peterson, E. Winston (Eds) Sign Language Interpreting and Interpreter Education: Directions for Research and Practice : Oxford University Press 2005 ISBN 9780195176940

[15] I. Stewart, W. McKee, “ ‘Talkshow’ : an investigation into the support of hearing impaired students by the use of continuous speech recognition software”, Proceedings of the Widening Access: International Perspectives Conference. 2001 Annexe, ISSN 1-903661-14-5

[16] (2001) Liberated Learning website [online] www.liberatedlearning.com accessed 11/4/2001

[17] K. Bain, S. Basson, M. Wald, “Speech Recognition in University Classrooms: Liberated Learning Project” , Proceedings of the fifth international ACM conference on Assistive technologies 2002 pp 192-196 ISBN:1-58113-464-9

[18] I.K. Jordan. The view from the president’s office Darby, A and Taylor, G. (Eds) (2003) Deaf Identities. Coleford : Douglas McLean. ISBN: 9780946252534

[19] A. Smith, Secretary of State for Work and Pensions, Written Statement in the House of Commons, House of Commons Hansard 18 March 2003

[20] A.S.Drigas, D.Kouremenos, S. Kouremenos and J. Vrettaros An e-Learning System for the Deaf people Proceedings of the ITHET 6th Annual International Conference 0-7803-9141-1

[21] B.S. Parton, Sign language recognition and translation: A Multidisciplined Approach From the Field of Artificial Intelligence, Journal of Deaf Studies and Deaf Education Volume 11 no 1, pp94- 101 Winter 2006 Online ISSN 1465-7325 - Print ISSN 1081-4159

[22] I. Stewart, M. Allan, D. Harrison “An investigation into voice recognition based assistive technology in higher education” Proceedings of the 4th European Conference on e-Learning 2005 pp 449-457 ISBN:1-905305-12-5

[23] A. Park, T.J. Hazen, J.R. Glass, J.R. “Automatic Processing of Audio Lectures for Information Retrieval: Vocabulary Selection and Language Modelling”, 2005. (ICASSP 2005). Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume: 1, pp 497- 500 ISSN: 1520-6149 ISBN: 0-7803-8874-7

[24] J. Glass, T. Hazen, L. Hetherington, C. Wang, "Analysis and processing of lecture audio data: Preliminary investigations", Proceedings of the HLT-NAACL 2004 Workshop on Interdisciplinary Approaches to Speech Indexing and Retrieval, pp. 9-12, Boston, Massachusetts, May, 2004.

[25] M. Wald, “An exploration of the potential of Automatic Speech Recognition to assist and enable receptive communication in higher education”, ALT-J, Research in Learning Technology, Vol. 14, No. 1, March 2006, pp. 9–20 ISSN 0968-7769 (print)/ISSN 1741-1629 (online)

[26] T. Mayes. “Flexible Delivery”, QAA enhancement theme. QAA 2006 pp5-13 ISBN1 84482 576 0

[27] (2008) The ECHO360 website [online], http://www.apreso.com/ Accessed 12/4/2008

[28] (2008) The Sonic Foundry website [online] http://www.sonicfoundry.com/ Accessed 13/4/08

[29] A. McClure, Lecture Capture: A Fresh Look, University Business, Professional Media Group, April 2008, online at http://www.universitybusiness.com/

[30] M. L Barbier, A Piolat,. L1 and L2 cognitive effort of notetaking and writing. In L. Alla, & J. Dolz (Eds.). Proceedings at the SIG Writing conference 2004, Geneva, Switzerland.

[31] M. Wald, "A Research Agenda for Transforming Pedagogy and Enhancing Inclusive Learning through Synchronised Multimedia Captioned Using Speech Recognition" in Proceedings of ED-MEDIA 2007: World Conference on Educational Multimedia, Hypermedia & Telecommunication, 2007. pp. 4479-4485

[32] Y. Rui, A. Gupta, J. Grudin, L. He, Automating lecture capture and broadcast: technology and videography, Journal of Multimedia Systems Volume 10, Number 1 / June, 2004, pp 3-15 Springer Berlin / Heidelberg ISSN 0942-4962 (Print) 1432-1882 (Online)

[33] M. Dimitrova, C. Sadler, S. Hatzipanagos and A. Murphy, Addressing Learner Diversity by Promoting Flexibility in e-Learning Environments, Proceedings of the 14th International Workshop on Database and Expert Systems Applications (DEXA 2003) ISSN 1529-4188/03S.

[34] D. Andone, J. Dron, L. Pemberton and C. Boyne, “The desires of digital students”, Proceedings of the 14th Association for Learning Technology International Conference pp189-200 ISBN 978-0-954587

Research on Disparity and Background-based Object Segmentation in Stereo Video

Tian-xiao Guan, Shi-gang Wang*, He-xin Chen, Jian Wei, Li-wei Chen
School of Communication Engineering
Jilin University, P.R.China
Nanhu Road 5372#, Changchun, China

Abstract-**In this paper, a new algorithm to segment object from a pair of stereo images is presented. First, we matched every pixel in corresponding scan line pairs using a one-dimensional window, obtained the disparity map and initial object segmentation mask. Second, constructed complete and reliable background region from the accumulated frames by using background statistical technology. As compared with current frame to segmented the accurate target object mask. Then a post-processing was applied to remove noise and smooth the object boundary of segmentation mask, thus we obtained the object. The algorithm does not need to know the shape and number of motion objects beforehand, and it can separate objects from stationary and complex background. The experiment shows that this algorithm has certain practicality.**

I. Introduction

Stereo vision infers three-dimensional (3-D) scene by generated recording two images with slightly different view points of the same scene, it provides a direct way of inferring the depth information by using the stereo pairs. Compare with the normal vision, stereo vision has more quantities of data, in order to save the storage and bandwidth, efficient coding techniques should be employed to reduce the data rate. Block-based and Object-based methods are often used to coding stereoscopic image sequences. Block-based approach is simple and easy to achieve and allowing more straightforward hardware implementations, but the subjective quality of reconstructed images may be bad at low bit-rates. Object-based schemes alleviate the problem of annoying coding errors, providing a more natural representation of the scene, excellent performance and by producing fewer annoying effects such as blocking artifacts and mosquito effects. Furthermore, important image areas such as facial details in face-to-face communications can be reconstructed with a higher image quality than with block-oriented hybrid coding. In addition, the ability of object-based coding techniques to describe a scene in a structural way, in contrast to traditional waveform-based coding techniques, opens new areas of applications, but they require a complex analysis phase to segment the scene into objects and estimate their motion and structure[1].

It is well known that video object segmentation is a crucial step in object-based stereo video coding. It is also an important step in many computer vision and multimedia tasks, including video editing, compositing, and the combination of real with synthetic video content. So object segmentation from images has long been an active area of research.

Disparity refers to the displacement between the corresponding points of a real-world point in the left and right images. We can also derive the depth information of the real-world point from the camera parameter and the disparity of the corresponding points. Using the depth information, the scene may be separated in layers and depth keying is possible. Additionally, accurate 3-D modeling of the scene structure may be achieved. The disparity estimation problem is similar to the motion estimation problem in that it requires to establish the correspondence between pixels in two images. For every point in one image, the task of disparity estimation is to determine its corresponding point in the other image, or mark it as being occluded in the other image. Disparity estimation algorithms can be generally grouped into content-based, feature-based and phase-based[2].

In this paper, we do histograms matching two times in order to reduce the influence of intensity differences between left and right image. First we match right histogram as left one is a standard, second match left one as right is a standard. Then we sub-sampling the images to decrease quantities of computing, match the individual pixels in one partition of the horizontal scan line in one image with the pixels in the corresponding partition of the scan line in the other image. We use a one-dimensional window to do the matching procedure[3]. After computing a disparity map, information between scan lines is used to refine the disparity map. According to the coherence principle, image areas with smooth disparity variation belong to the same physical object. The regions which are discontinuity can let us get the initial object mask[4]. Then we over-sampling the mask to original size. The regions which are outside the object mask form the background. The whole background can be constructed from these regions of accumulated frames, compare the background with current frame, then we can segment the accurate target object mask, process the mask image using morphological methods, obtain the accurate object.

The rest of the paper is organized as follows: In section 2, we describe the popular camera configuration in stereoscopic systems. The disparity estimate procedure is introduced in section 3. Section 4 discusses the background constructed method algorithm. The experimental results are presented in section 5. Finally, we conclude with a discussion of further research directions in Section 6.

---

*Shi-gang Wang is contact author.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_54, © Springer Science+Business Media B.V. 2010

II. Camera Configuration In Stereoscopic Systems

Stereoscopic systems are designed to emulate human binocular imaging system: capturing a 3-D scene using two cameras located in slightly shifted positions. The two cameras contain significant depth information. There are mainly two kinds of stereoscopic systems: parallel camera configuration and converging camera configuration. Parallel camera configuration is the most popular. In this stereoscopic system, two cameras with parallel imaging planes are located on the same X-Y plane of the world coordinate. The parallel camera configuration is shown in Fig.1.

C_w denotes a chosen world coordinate. C_l and C_r denote the left and right camera coordinates respectively. z_l and z_r are the corresponding points of a real-world point in the left and right images. The distance between the two cameras is denoted by B, called the baseline distance.

If we set the original of the world coordinate halfway between the two cameras, and assume that the two cameras have the same focal lengths, denoted by F. Then we can get the following equations:

$$X_l = X + \frac{B}{2}\ ,\ X_r = X - \frac{B}{2}$$
$$Y_l = Y_r = Y\ ,Z_l = Z_r = Z \tag{1}$$

$$x_l = F\frac{X + B/2}{Z}, x_r = F\frac{X - B/2}{Z}, y_l = y_r = y = F\frac{Y}{Z} \tag{2}$$

The relationship between disparity vector Z and depth d is described as:

$$d = z_l - z_r = \frac{FB}{Z} \tag{3}$$

It is obvious that corresponding points in the left and right images are on the same horizontal scan line, and that the closer object has a larger horizontal disparity.

III. Disparity Estimate

A. Constraints in Stereo Image

Several suitable constraints can be applied in order to obtain an accurate solution. These constraints can be very helpful for us in disparity estimation.

Epipolar Constraint: The corresponding pixels in a stereo pair always lie on respective epipolar lines. For the parallel camera configuration, the epipolar lines are parallel to horizontal scan lines of the images, so that one can constrain the search with in the horizontal line on which the right image point is located. This is an important fundamental theory of our paper.

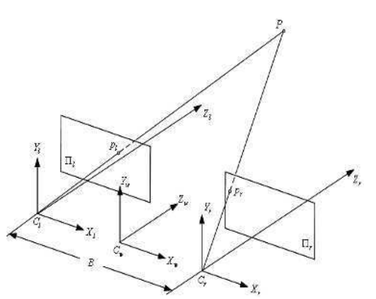


Fig.1. Parallel camera configuration

Uniqueness Constraint: A sampling position in one image must be the projection of only one real-world surface point because each point has a unique physical position.

Ordering Constraint: Assuming opacity of the imaged objects, the relative positions of points in an object are the same in both views of the scene: a feature point to the left of another in the left view lies to the left in the right view as well.

B. Disparity Estimation

Because of the angle of incidence, the image intensity has a little difference between left and right image from the same stereo sequence, it will make the disparity estimate not quite accurate. Histograms matching between left and right image can decrease the influences of light. In this paper, we do histograms matching two times in order to get a better result, first we match right histogram as left one is a standard, second match left one as right is a standard. Then we sub-sampling the images to decrease quantities of computing.

The image intensity at the point $z_l = (x_l, y_l)$ in the left image is $I_l(z)$. Two points $z_l = (x_l, y_l)$ and $z_r = (x_r, y_r)$ in the left and right images should be matched if and only if they are image plane projections of the same real-world surface point. In this case, the disparity vector pointing from z_l to z_r is given by:

$$d_l(z_l) = (x_r - x_l, y_r - y_l) \tag{4}$$

In standard stereoscopic images, the relevant depth information is contained in the horizontal component of the disparity field, so we can just consider the horizontal component x of the vector z. Therefore, $d_l(z_l)$ can be simplified as

$$d_l(z_l) = x_r - x_l \tag{5}$$

We use a one-dimensional window to do the matching procedure, which has a width of $2\omega+1$ and centered on the pixel $z_l(x_l, y_l)$ in the left image, compute the minimum absolute difference (MAD) with the $2\omega+1$ pixels which are in (δ_1, δ_2) on the corresponding horizontal scan line in right image. The criterion MAD is:

$$M = \sum_{i=-\omega}^{\omega} \left| x_{l+i} - x_{r+i} \right| \tag{6}$$

Search all pixels in (δ_1, δ_2), record coordinates $z_r(x_r, y_r)$ which makes M minimum. $z_r(x_r, y_r)$ is corresponding pixel of $z_l(x_l, y_l)$, $d_l(z_l) = x_r - x_l$ is disparity between $z_l(x_l, y_l)$ and $z_r(x_r, y_r)$. Pixel has the same coordinates as $z_l(x_l, y_l)$ is given value of $d_l(z_l) = x_r - x_l$ in left disparity map.

The matching process is performed bidirectionally for all image points in each scan line. After local matching, a dense disparity map is available.

In order to reduce noise in the disparity map, a median filter is adopted. Median filter is one of non-linear filters, it has simple operation and can avoid illegibility of edges in images.

According to the coherence principle, image areas with smooth disparity variation belong to the same physical object, the regions which are discontinuity can let us get the object mask. We compute the average of disparity value in a $n_1 \times n_2$ region which is included in the target object region using the equation:

$$\overline{d} = \frac{1}{n_1 n_2} \sum_{z_i \in (n_1 \times n_2)} d(z_i) \tag{7}$$

Object region is obtained by the equation below:

$$Object(z_i) = \begin{cases} 1, & \left|d(z_i) - \overline{d}\right| \le \delta \\ 0, & \left|d(z_i) - \overline{d}\right| > \delta \end{cases} \tag{8}$$

δ is a number given by experiment data. The region which makes $Object(z_i) = 1$ is object region, and which makes $Object(z_i) = 0$ is background region. The object region is in the disparity level between $\overline{d} - \delta$ and $\overline{d} + \delta$. Then we get the initial object mask by disparity estimating.

IV. Background Statistical Technology

The characteristics of motion regions are inscrutable, accurate figure information of objects cannot be constructed from the image, but background information is easy to constructed without the characteristics. In this paper we segment objects based on background information. Initial stereo object mask is get by disparity estimating, as we know the regions outside the mask can construct backgrounds, compute the correlation between object masks of some frames in statistics, find out the frames which make the correlation least, combine the regions which are not include in object masks of these frames, then we can construct whole backgrounds.

The difference between the background image and current frame called background difference, we obtained the object segmentation mask by changing the difference image to binary image. It is ineluctable that the object segmentation mask has noise in it because of the noise caused by camera and motion of objects. Then we process the mask image using morphological methods, thus we can obtain an accurate object segmentation mask and extract target object. The algorithm does not need to know the shape and number of motion objects beforehand, and it can separate objects from stationary and complex background.

The background statistical scheme is shown in Fig.2. The scheme of segmentation based on disparity estimate and background statistical is shown in Fig.3.

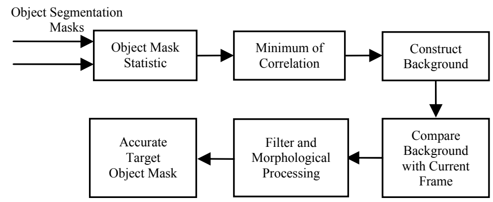


Fig.2. Background statistical scheme

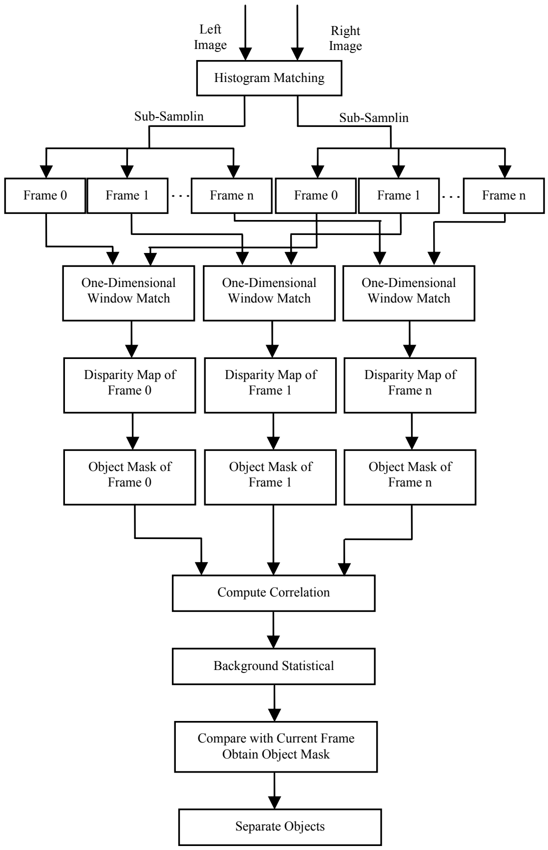
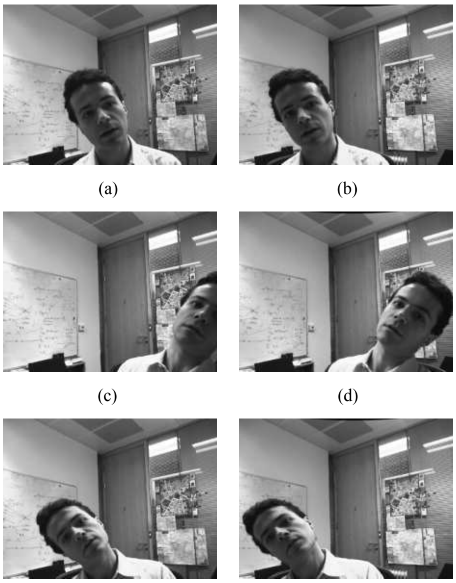


Fig.3. segmentation based on disparity estimate and background statistical scheme

V. Experimental Results

We segment a rest stereoscopic image sequences to checkout the algorithm used in this paper. The stereoscopic image sequences have an indoor, stationary and complex background, and have a little difference of intensity between left and right image, with the size of 320*240. We match histograms two times, and sub-sampling the images, then obtain disparity map by one-dimensional window matching. Adopt median filter to decrease the non-linear noise, segment object according to the coherence principle, process the initial mask image using morphological methods. Find out the frames which make the correlation minimum, do background estimation to get the whole background image, compare with the current frame to compute the accurate object mask, finally separate the objects.

Fig.4 shows 60th, 110th, 136th frames of the original stereo image sequence. The constructed backgrounds are shown in Fig.5. After using the refinement algorithm, we can get more accurate results, as shown in Fig.6.

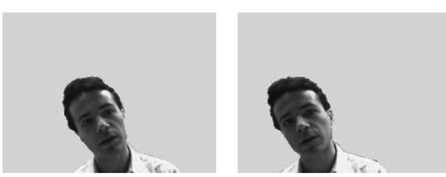


(a) (b) (c) (d) (e) (f)

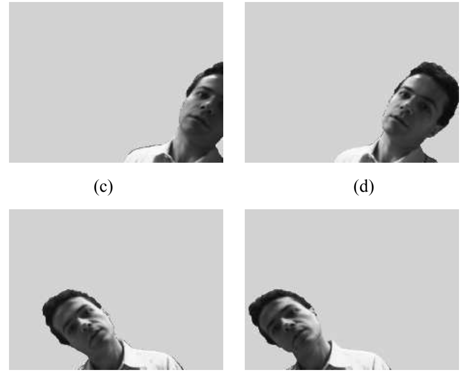
Fig.4. (a) 60th frame of left image (b) 60th frame of right image (c) 110th frame of left image (d) 110th frame of right image (e) 136th frame of left image (f) 136th frame of right image

(a) (b)

Fig.5. (a) Left background (b) Right background



(a) (b)



(c) (d) (e) (f)

Fig.6. (a) 60th object of left image (b) 60th object of right image (c) 110th object of left image (d) 110th object of right image (e) 136th object of left image (f) 136th object of right image

VI. CONCLUSIONS

In this paper, an algorithm to segment object based on disparity map and constructed backgrounds image is proposed. Experimental results have shown the performance of the proposed scheme to be better on condition that images have stationary and complex background even if have more than one object in it. But the constructed backgrounds have a great influence on performance. In future work, we anticipate that by improving the disparity estimation algorithm, it should be possible to get more accurate disparity map. By adding motion information in the segmentation procedure, it will be promising to get more satisfying results.

REFERENCES

[1] M. G. Strintzis, and S. Malassiotis, "Object-Based coding of stereoscopic and 3D Image Squences," *IEEE Signal Processing Magazine,* pp. 14-28, 1999.

[2] F. Edouard, and C. Bertrand, "Depth-based segmentation," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 7, pp.237-240, January 1997.

[3] Qian Zhou, Shi-gang Wang, Hong Yang. Research on Disparity-based Object Segmentation in Stereo. Proceeding of the Asia-Pacific Workshop on Visual Information Processing (VIP06),2006:115-118.

[4] D. Scharstein, and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision* vol. 47, pp. 7-42, January 2002.

[5] S. Birchfield, and C, Tomasi, "Depth discontinuities by pixel-to-pixel stereo," *International Journal of Computer Vision,* vol. 35, pp.269-293, March 1999.

[6] W. Woo, and A. Ortega, "Overlapped block disparity compensation with adaptive windows for stereo image coding," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 10, pp. 194-200, February 2000.

[7] D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Object-Based Coding of Stereo Image Sequences Using Joint 3-D Motion/Disparity Compensation," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 7, pp.312-327, April 1997.

[8] C. R. Wang, and G. H. Gu, "A video object segmentation algorithm with background statistical technology," *Opto-Electronic Engineering,* vol. 31, pp. 57-60, August 2004.

Dynamic Modeling and VR Simulation of 3DOF Medical Parallel Robots

Nadia Ramona Raţ[1], Mircea Neagoe[1], Dorin Diaconescu[1], Sergiu Dan Stan[2], Radu Bălan[2]
[1] University 'Transilvania' of Brasov, Department of Product Design and Robotics
[2] Technical University of Cluj-Napoca, Department of Mechatronics
E-mail: ncretescu@unitbv.ro, mneagoe@unitbv.ro, dvdiaconescu@unitbv.ro, sergiustan@ieee.org, radubalan@yahoo.com

Abstract-**The paper presents dynamic modeling and VR (virtual reality) simulation issues for two 3 DOF medical parallel robots with three coupled motions (Orthoglide robot and with one coupled motion, robot of type 2PRRR+1PRPaR). Kinematical and dynamical models, followed by a VR application with control aspects are presented for these robots. The innovative user interface for high-level control of the two parallel robots, presented in the paper, was developed in MATLAB - Simulink and SimMechanics environment. This kind of parallel robots can be successfully applied for medical applications where accuracy and high dynamic behavior are required. This research will lay a good foundation for the development of medical parallel robots.**

I. Introduction

Over the last couple of decades parallel robots have been increasingly studied and developed from both theoretical viewpoint as well as for practical applications. Advances in computer technology and development of sophisticated control techniques have allowed for more recent practical implementation of parallel manipulators. Some of the advantages offered by parallel manipulators, when properly designed, include an excellent load-to-weight ratio, high stiffness and positioning accuracy and good dynamic behavior ([2], [3]). The ever-increasing number of publications dedicated to parallel robots illustrated very well this trend in research.

The parallel robots are mechanisms with closed kinematics chains, composed by an end-effector (the mobile platform) with n degrees of freedom, connected to the fixed base by two or more kinematical chains called limbs or legs. A simple or a complex kinematics chain can be associated with each limb [2].

In this paper a kinematics and a dynamic modeling is presented for two parallel robots with different degrees of coupling: one and three, respectively.

Different methods can be applied to dynamic modeling of parallel robots. In this paper, the Lagrange method with multipliers is used to derive the closed form dynamic model in the hypothesis of rigid links. Based on a representative numerical example, these results are verified using the CAD model implemented in ADAMS simulation environment.

The Virtual Reality (VR) immerses the user in a three-dimensional (3D) environment that can be actively interacted and explored. Virtual reality environment tool is used by many researchers in design, development and manufacturing of the robotic industry [4].

Using the virtual reality simulation with a virtual robot, a three dimensional design and the real-time behavior of the robot can be observed; that fact is relatively new and allows testing the robot before accomplishment a physical implementation. In this way, resources (money and time) can be saved and various problems can be solved from the design stage.

This paper presents the necessary steps for developing the virtual environment for kinematic simulation, starting from the SolidWorks model of these two parallel robots.

II. 3PRPaR and 2PRRR+1PRPaR Parallel Robots Description

The first parallel robot studied is Orthoglide, a 3 DOF parallel robot ([1], [5]) with three coupled motions of type 3PRPaR (3 – Prismatic – Revolute – Parallelogram - Revolute), composed by a mobile platform connected to the fixed frame by three kinematics chains (Fig. 1).

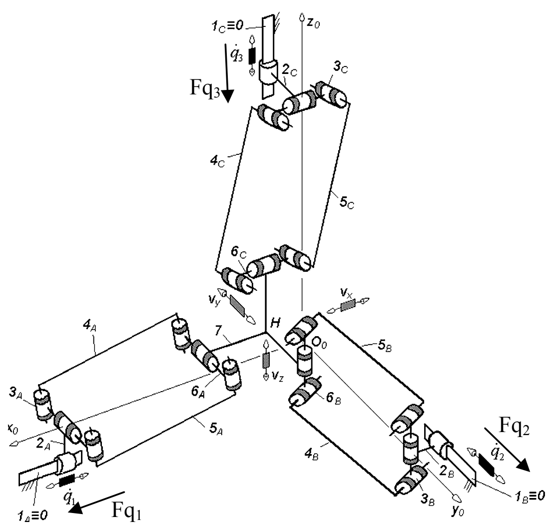


Fig. 1. Kinematic structure of Orthoglide parallel robot also known as 3PRPaR parallel robot [1].

The second parallel robot studied is a 3DOF parallel robot [1], [2] with one coupled motions of type 2PRRR+1PRPaR (2 – Prismatic – Revolute – Revolute – Revolute and

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_55, © Springer Science+Business Media B.V. 2010

1 – Prismatic – Revolute – Parallelogram – Revolute), composed by a mobile platform connected to the fixed frame by three kinematics chains (Fig. 2).

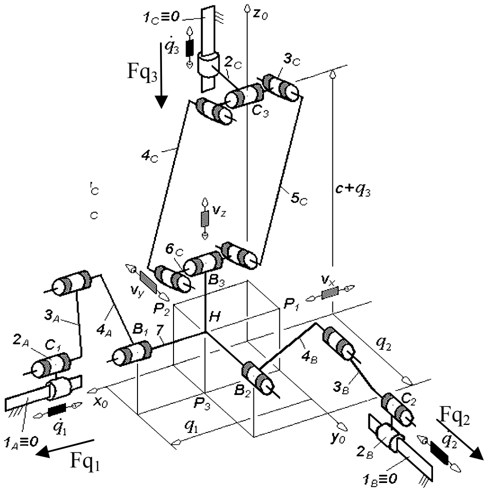


Fig. 2. Kinematic structure of parallel robot of type 2PRRR+1PRPaR [1]

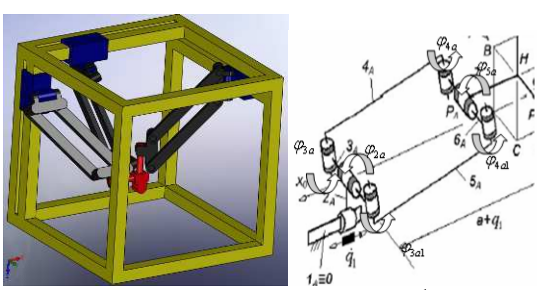
More information regarding the uncoupled parallel robots can be found in ref. [1] and [2].

III. Kinematics Modeling Of the 3 DOF Parallel Robots and Virtual Reality Tool

These parts of paper present the kinematics model for the two parallel robots. Typically, the study of the robot kinematics is divided into two parts, inverse kinematics and forward (or direct) kinematics.

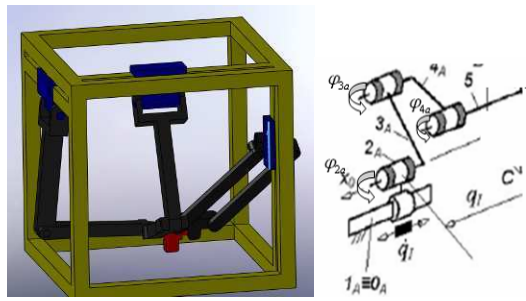
The inverse kinematics problem involves a known pose (position and orientation) of the output platform of the parallel robot to a set of input joint variables that will achieve that pose.

The forward kinematics problem involves the mapping from a known set of input joint variables to a pose of the moving platform that results from those given inputs. However, the inverse and forward kinematics problems of our parallel robots can be described in closed form.



a b

Fig. 3. The CAD model of Orthoglide parallel robot (a) and the notations for passive joints.



a b

Fig. 4. The CAD model of parallel robot de type 2PRRR+1PRPaR (a) and the notations for passive joints (for the uncoupled motion links)

A. Direct and inverse kinematics model of Orthoglide parallel robot and of 2PRRR+1PRPaR parallel robot

The direct kinematical model (Eq. 1) determines the mobile platform velocity (V_{xh}, V_{yh}, V_{zh}) in function of drivers' motion ($\dot{q}_1, \dot{q}_2, \dot{q}_3$).

$$\begin{bmatrix} v_{xh} \\ v_{yh} \\ v_{zh} \end{bmatrix} = J_h \cdot \begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \end{bmatrix}, \qquad (1)$$

where: J_h is the Jacobian of the system.

The inverse model (geometrical and kinematical) permit to obtain the liaison between $\{q_1, q_2, q_3\}$ and $\{x_h, y_h, z_h\}$ - the inverse geometrical model – eq. 3 for 2PRRR+1PRPaR parallel robot and eq. 2 for Orthoglide parallel robot (notations presented in fig. 1-4).

$$\begin{cases} q_1 = -\cos\left(2\arctan\left(\dfrac{z_h}{\sqrt{\dfrac{l_{4a}^2 - y_h^2}{l_{4a}^2}}\, l_{4a}\left(1+\sqrt{1-\dfrac{z_h^2}{l_{4a}^2 - y_h^2}}\right)}\right)\right)\cdot \cos\left(2\arctan\left(\dfrac{y_h}{l_{4a}\left(1-\sqrt{1-\dfrac{y_h^2}{l_{4a}^2}}\right)}\right)\right)\cdot l_{4a} + l_{7a} + x_h \\ q_2 = -\cos\left(2\arctan\left(\dfrac{x_h}{\sqrt{\dfrac{l_{4b}^2 - z_h^2}{l_{4b}^2}}\, l_{4a}\left(1+\sqrt{1-\dfrac{x_h^2}{l_{4b}^2 - z_h^2}}\right)}\right)\right)\cdot \cos\left(2\arctan\left(\dfrac{z_h}{l_{4b}\left(1-\sqrt{1-\dfrac{z_h^2}{l_{4b}^2}}\right)}\right)\right)\cdot l_{4b} + l_{7b} + y_h \\ q_3 = -\cos\left(2\arctan\left(\dfrac{y_h}{\sqrt{\dfrac{l_{4c}^2 - x_h^2}{l_{4c}^2}}\, l_{4c}\left(1+\sqrt{1-\dfrac{y_h^2}{l_{4c}^2 - x_h^2}}\right)}\right)\right)\cdot \cos\left(2\arctan\left(\dfrac{x_h}{l_{4c}\left(1-\sqrt{1-\dfrac{x_h^2}{l_{4c}^2}}\right)}\right)\right)\cdot l_{4c} + l_{7c} + z_h \end{cases}$$

(2)

$$
\begin{aligned}
q_1 &= x_h - l_{7a} \\
q_2 &= y_h - l_{7b} \qquad (3) \\
q_3 &= -\cos\left(2\arctan\left(\frac{y_h}{\sqrt{\frac{l_{4c}^2 - x_h^2}{l_{4c}^2}}\, l_{4c}\left(1+\sqrt{1-\frac{y_h^2}{l_{4c}^2 - x_h^2}}\right)}\right)\right)\cdot\cos\left(2\arctan\left(\frac{x_h}{l_{4c}\left(1-\sqrt{1-\frac{x_h^2}{l_{4c}^2}}\right)}\right)\right)\cdot l_{4c} + \\
&+ l_{7c} + z_h
\end{aligned}
$$

Deriving the rel. 2 and 3 the inverse kinematical model (eq. 4) is obtained:

$$
\begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \end{bmatrix} = J_h^{-1} \cdot \begin{bmatrix} v_{xh} \\ v_{yh} \\ v_{zh} \end{bmatrix}. \qquad (4)
$$

B. Virtual Reality tools of parallel robots

An innovative user interface for high-level control of robot manipulators is presented in this section. The interface is based on a virtual reality approach in order to provide the user with an interactive 3D graphical representation of the parallel robot. The interface was designed to give to a novice user an intuitive tool to control any kind of mechanical structure (serial, parallel or hybrid), requiring no programming skills. Computer based simulation allows mimicking of a real life or potential situations. SimMechanics models, however, can be interfaced seamlessly with ordinary Simulink block diagrams. For example, this enables user to design the mechanical and the control system in one common environment.

The actuators and the control algorithm were modeled with Simulink. The dynamic model of the mechanical structure was imported from SolidWorks using SimMechanics from MATLAB/Simulink (Fig. 5).

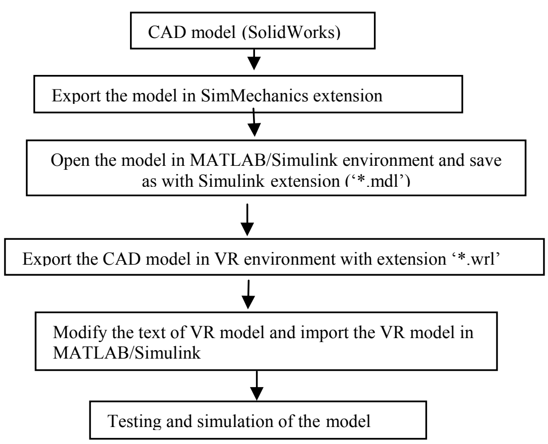


Fig. 5. The algorithm for the VR model obtained from the SolidWorks CAD model of the robot

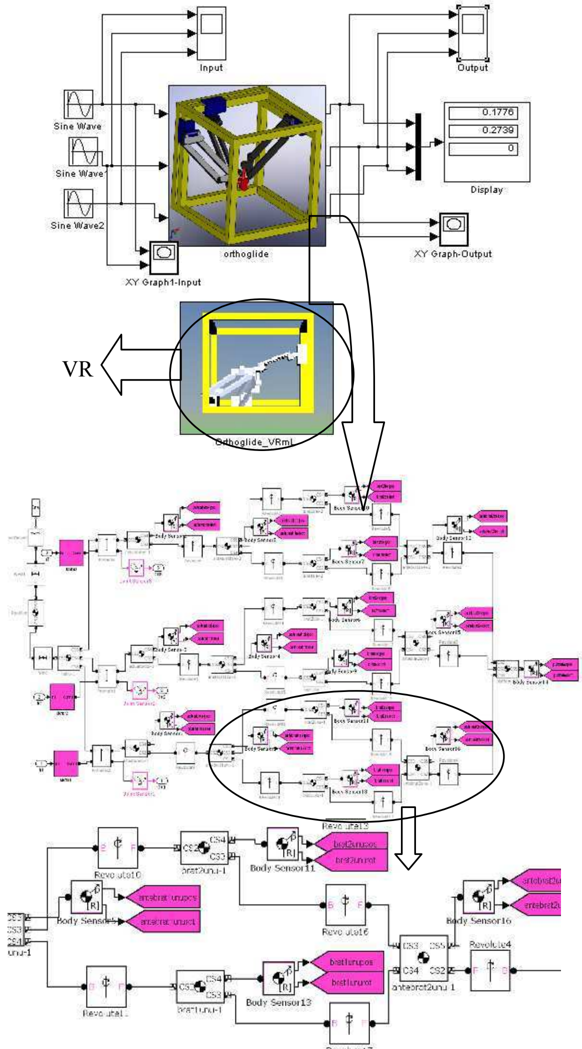


Fig. 6. The complete model for Orthoglide parallel robot with VR

In figure 6 the complete model with VR (developed in fig7) of Orthoglide parallel robot is presented.

Virtual Reality Toolbox for MATLAB makes more realistic renderings of bodies possible. Arbitrary virtual worlds can be designed with Virtual Reality Modeling Language (VRML), and interfaced to the SimMechanics model. The user simply describes the geometric properties of the robot first. Then, in order to move any part of the robot through 3D input devices, the inverse kinematics is automatically calculated in real time. The interface was also designed to provide the user decision capabilities when problems such as singularities are encountered.

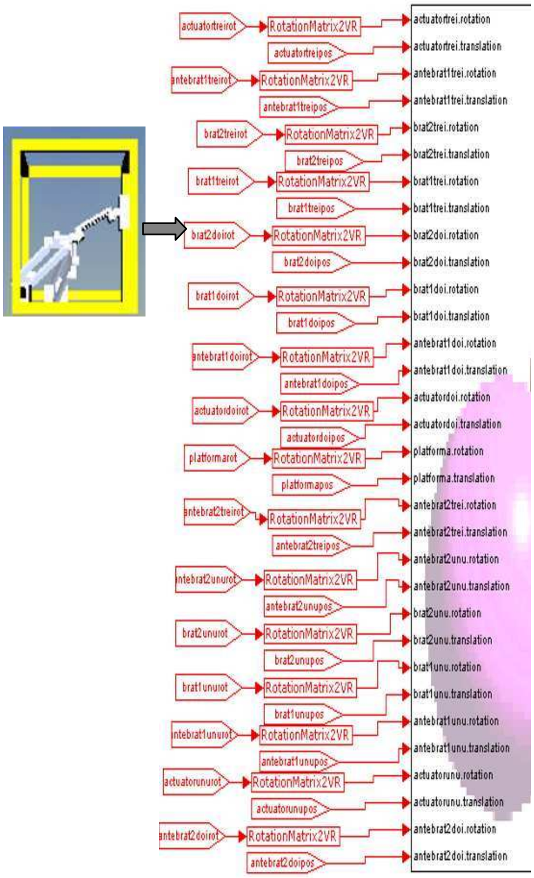


Fig. 7 Virtual Reality tool for Orthoglide parallel robot

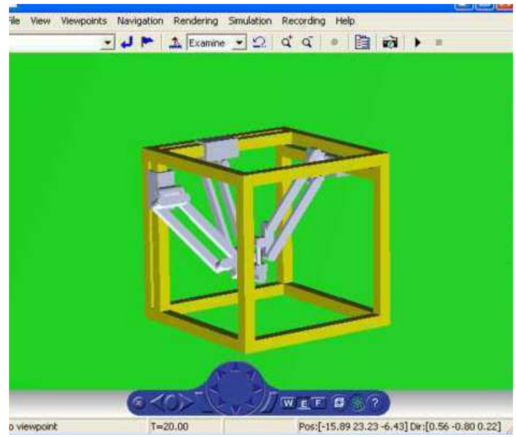


Fig. 8. Virtual Reality model of parallel robot de type Orthoglide

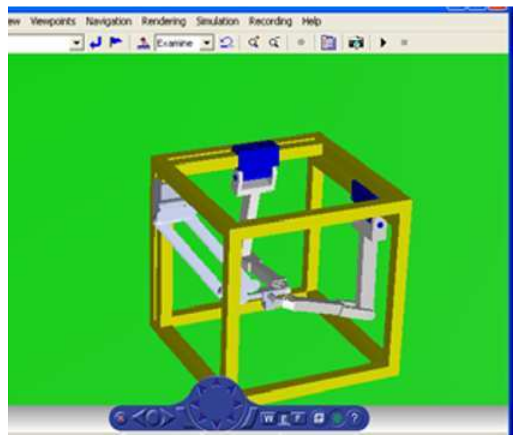


Fig. 9. Virtual Reality model of parallel robot de type 2PRRR+1PRPaR

VR interface enables users to interact with the robot in an intuitive way. This means that the operator can pick and choose any part of the robot and move it using translation and rotation using 3D sensors, as easily as a "drag and drop" operation is. Thus, trajectories can be easily defined, optimized, and stored. Not only that, but the virtual world can be accessed and controlled via the Internet, too.

The use of a VR interface to simulate robots drastically improves the "feeling" for the robot.

In particular, the interface allows user to understand the behavior of an existing robot, and to investigate the performance of a newly designed structures without the need and the cost associated with the hardware implementation.

SimMechanics offers the possibility to visualize and animate the robot. The visualization tool can also be used to animate the motion of the system during simulation. The bodies of the robot can be represented as convex hulls (Fig. 8-9).

IV. Dynamics Modeling of the Parallel Robots

The dynamic modelling has made using the Lagrange multipliers method in the rigid link hypothesis, where the gravity acceleration is considering on Z axe in negative sense.

Lagrange equation with multipliers is described by:

$$\sum_{i=1}^{k}\lambda_i \frac{\partial \Gamma_i}{\partial q_j} = \frac{d}{dt}\left(\frac{\partial L}{\partial q_j}\right) - \frac{\partial L}{\partial q_j} - \hat{Q}_j \ , \qquad (5)$$

where: λ_i - Lagrange multipliers; L - Lagrangian of the system; q_j - displacement of the actuators; $\hat{Q}_j$ - generalizing external force.

The Lagrangian of the system is done:

$$L = \sum_{i=1}^{n} E_{ci} + \sum_{i=1}^{n} E_{pi} \ , \qquad (6)$$

where: E_{ci} - kinetic energy; E_{pi} - potential energy;

The next step is to identify equations set for introduction of the multipliers (Γ_i):

- for Orthoglide parallel robot:

$$\begin{aligned}
Ecu1 &= \sin(phi3a)\, l4a - q2 - \cos(phi2b)\cos(phi3b)\, l4b + l7b \\
Ecu2 &= -\frac{1}{2} l4a \sin(phi2a + phi3a) - \frac{1}{2} l4a \sin(phi2a - phi3a) - q3 - \cos(phi2c)\cos(phi3c)\, l4c + l7c \\
Ecu3 &= -\sin(phi2b)\cos(phi3b)\, l4b - q1 - \frac{1}{2} l4a \cos(phi2a - phi3a) - \frac{1}{2} l4a \cos(phi2a + phi3a) + l7a \\
Ecu4 &= \sin(phi3b)\, l4b - q3 - \cos(phi2c)\cos(phi3c)\, l4c + l7c \\
Ecu5 &= \sin(phi3c)\, l4c - q1 - \frac{1}{2} l4a \cos(phi2a - phi3a) - \frac{1}{2} l4a \cos(phi2a + phi3a) + l7a \\
Ecu6 &= -\sin(phi2c)\cos(phi3c)\, l4c - q2 - \cos(phi2b)\cos(phi3b)\, l4b + l7b
\end{aligned} \tag{7}$$

- for 2PRRR+1PRPaR parallel robot:

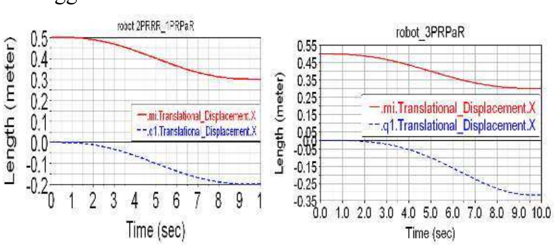
$$\begin{aligned}
Ecu1 &= \cos(phi2a)\, l3a + l4a \cos(phi2a + phi3a) - q2 + l7b \\
Ecu2 &= \sin(phi2a)\, l3a + l4a \sin(phi2a + phi3a) - q3 - \cos(phi2c)\cos(phi3c)\, l4c + l7c \\
Ecu3 &= \sin(phi2b)\, l3b + l4b \sin(phi2b + phi3b) - q1 + l7a \\
Ecu4 &= \cos(phi2b)\, l3b + l4b \cos(phi2b + phi3b) - q3 - \cos(phi2c)\cos(phi3c)\, l4c + l7c \\
Ecu5 &= \sin(phi3c)\, l4c - q1 + l7a \\
Ecu6 &= -\sin(phi2c)\cos(phi3c)\, l4c - q2 + l7b
\end{aligned} \tag{8}$$

Finally, the analytical expression of the driver forces Fq_1, Fq_2 and Fq_3 (see Fig. 1 and 2) are obtained.

Finally, the analytical expression of the driver forces Fq1, Fq2 and Fq3 (see Fig. 1 and 2) are obtained. Even the analytical dynamic model is complex; the method used can be applied to derive the dynamical model of any parallel robots with 3 DOF and can be extending for 4 and more DOF parallel robots.

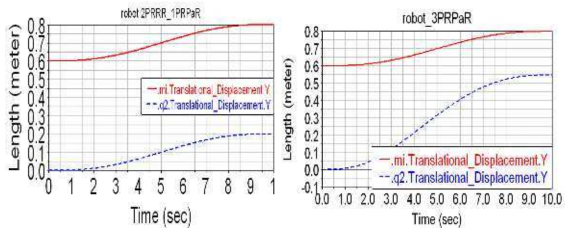
V. Simulation Results

For testing the obtained models, it has been chosen a trajectory between two points in Cartesian space after polynomial low of fifth degree order. Regarding the displacements, it can be seen that for the same desired displacement of the end-effector for both parallel robots, the required displacements in motor joints in case of the coupled motions (for Orthoglide and for q_3 from 2PRRR+1PRPaR) are bigger.



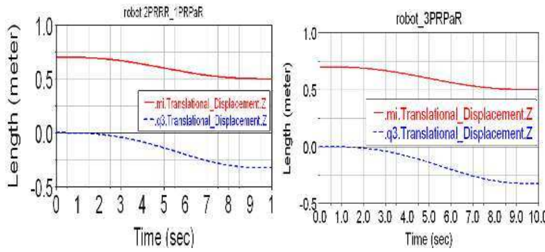
a b

Fig. 10. Displacement after X axis for the end-effectors (red-continues line) and for q1 (blue – dashed line) of parallel robot de type 2PRRR+1PRPaR (a) and Orthoglide (b)



a b

Fig. 11. Displacement after Y axis for the end-effectors (red-continues line) and for q1 (blue – dashed line) of parallel robot de type 2PRRR+1PRPaR (a) and Orthoglide (b)

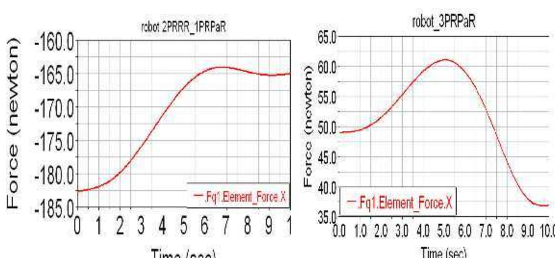


a b

Fig. 12. Displacement after Z axis for the end-effectors (red-continues line) and for q1 (blue – dashed line) of parallel robot de type 2PRRR+1PRPaR (a) and Orthoglide (b)

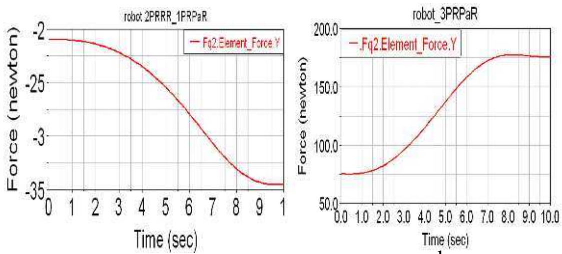
From the kinematic point of view it was noticed first a motion decoupling in the case of the parallel robot with 2 decoupled motions and on the other hand the influence of the motion coupling on the kinematic behavior. In the followings it will be studied also the influence of these coupling and decoupling on the developed forces in the motor joints.

It has to be mentioned that for studying comparatively these motor forces a total mass equivalent has been taken into consideration for both parallel robots. It can be seen that in both cases the maximum effort is for motor force Fq3 (Fig. 15) due to the coupling from motion after z-axis and gravity acceleration.



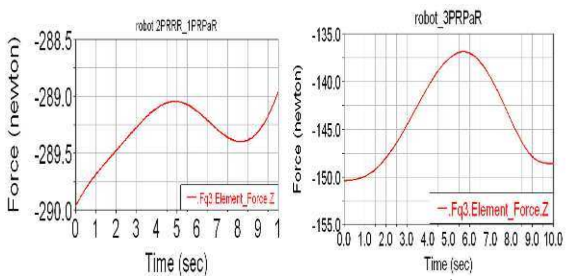
a b

Fig. 13. Driver forces Fq1 of parallel robot de type 2PRRR+1PRPaR (a) and Orthoglide (b)



a b

Fig. 14. Driver forces Fq2 of parallel robot de type 2PRRR+1PRPaR (a) and Orthoglide (b)



a b

Fig. 15. Driver forces Fq3 of parallel robot de type 2PRRR+1PRPaR (a) and Orthoglide (b)

This simulation results was done for the analytical dynamic model. Finally, a simulation result is presented for the complete model with VR of Orthoglide parallel robot presented in figure 16 for a circle in XY plane.

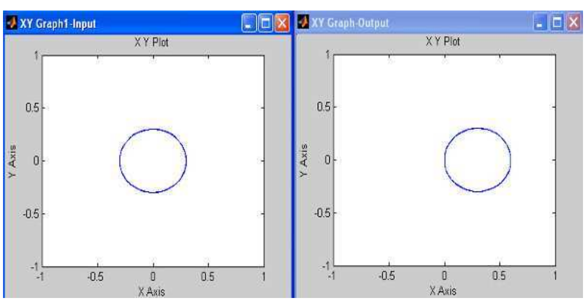


Fig.16 Simulation results for Orthoglide with VR

CONCLUSIONS

The paper presents a novel Virtual Reality Interface for the 3 DOF medical parallel robots control. An evaluation model from the Matlab/SimMechanics environment was used for the simulation. An interactive tool for dynamics system modeling and analysis was presented and exemplified on the control in Virtual Reality environment of these medical parallel robots.

A dynamic modeling of maximally regular parallel robots [2], with application to the coupled topology has been presented in this paper. A closed form dynamic model has been set up in the rigid link hypothesis. The numerical simulations on a given trajectory emphasize the influence on the driving forces. With obtained analytical dynamic model for both parallel robots a control model can be add to complete the developed study.

The main advantages of this parallel manipulator are that all of the actuators can be attached directly to the base, that closed-form solutions are available for the forward and inverse kinematics, and that the moving platform maintains the same orientation throughout the entire workspace. By means of SimMechanics, the robotic system in represented as a block of functional diagrams. Besides, such software packages allow visualizing the motion of mechanical system in 3D virtual space. Especially non-experts will benefit from the proposed visualization tools, as they facilitate the modeling and the interpretation of results. The research presented will lay a good foundation for the development of medical parallel robots.

ACKNOWLEDGMENT

This work was supported by the Ministry of Education from Romania and Research grant PARTENERIATE no. 3280 entitled ‘Complex mechatronics systems for medical applications’.

REFERENCES

[1] Gogu, G., “Evolutionary morphology: a structured approach to inventive engineering design”, Invited paper, *Proceedings of the 5th International Conference on Integrated Design and Manufacturing in mechanical Engineering*, Bath, 5-7 April 2004.

[2] Gogu, G., “Structural synthesis of parallel robots”, *Part 1: Methodology*, Springer Verlag, 2008.

[3] Halliday, S., Green, M., (1994) A Geometric Modeling and Animation System for Virtual Reality, *Virtual Reality Software and Technology (VRST 94)*, 71-84, Singapore, August 1994.

[4] Stan, S.-D, Manic, M., Mătieș, M., Bălan, R., “Kinematics Analysis, Design, and Control of an Isoglide3 Parallel Robot (IG3PR)”, *IECON 2008*, The 34th Annual Conference of the IEEE Industrial Electronics Society, Orlando, USA, November 10-13, 2008.

[5] Wenger, P.; Chablat, D., ”Kinematic analysis of a new parallel machine tool: the Orthoglide”. *In: Lenarčič J, Stanišić ML (eds) Advances in Robot Kinematics*. Kluwer Academic Publishers, pp 305-314.

Possibilities Concerning Control of a Laser Welding Cell Utilizing Servomotors and PLC

R. Pacurari, M. Crenganis, A. Csiszar, C. Brisan
Technical University of Cluj Napoca
Muncii str. 105
Cluj Napoca, 400641, Romania

***Abstract*-This paper presents laser welding cell. The mechanical structure is based on the Partner robot, it includes Wittenstein Ternary servo motors and it is controlled by a Schneider Twido PLC. It represents a reconfigurable manufacturing system.**

I. Introduction

The propose of this paper is to present a possible architecture of a complete laser welding cell, based on a parallel robot, moved by servo motors, equipped with conveyor belts. The whole architecture is controlled by a programmable logic controller (PLC), and it is supervised by a PC.

II. Partner Parallel Robot

A. Parallel robots

A parallel robot is a mechanism with at least one closed loop, topologically formed from one fixed platform and at least one mobile platform. These platforms are connected by at least two open loops. Good dynamic behavior (high stiffness), a high accuracy and a good ratio between total mass and manipulated mass are just few advantages of parallel robots compared with serial type. A modular parallel robot consists into independent designed modules, which contain actuator(s), passive joints and links [4, 10, 13, 14, 17]. These modules can be connected into a given topology in order to obtain a large variety of parallel robots. The main advantages of modular design consist into great flexibility, ease of maintenance, short time for development of new variant etc.

B. Modular Synthesis

The structural synthesis of parallel mechanisms could be made if the relation of the number of degrees of freedom (dof) it is considered:

$$M = (6-m)\cdot n - \sum_{k=1}^{5}(k-m)\cdot C_k - M_p \qquad (1)$$

where m is the number of common restrictions for all elements, n is the number of the mobile elements, k is the number of restrictions which define a joint (for example in the case of prismatic joint k=5), Ck is the number of joints with (6-k) degrees of freedom and MP is the number of identical degrees of freedom.

In the case of parallel mechanisms without common restrictions and also without identical degrees of freedom the relation (1) it becomes:

$$M = 6\cdot n - \sum_{k=1}^{5} k\cdot C_k \qquad (2)$$

Let be N the number of mobile platforms and D_k – the number of joints with (6-k) degrees of freedom which directly connect the platforms of the mechanism. With these notations it results:

$$M = 6(n_1 + N) - \sum_{k=1}^{5} k\cdot (C_k + D_k) \qquad (3)$$

where n1 is the number of the elements which compose the loops which connect the platforms of the mechanism.

We can also assume (Fig.1) two types of basic modules (named basic legs) which can connect the platforms of the mechanism (f is number of degrees of freedom of the joint). Let a1 be the number of the loops with prismatic - universal - spherical (PUS) topology, let a2 be the number of the loops with prismatic - rotational - spherical (PRS) topology.

In the case of parallel mechanisms which are used in the field of machine tools, it is common to consider:

$$N = 1, D_k = 0, k = \{1,...,5\} \qquad (4)$$

With these notations, the relation (3) becomes:

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_56,

$$M = 6 - a_2 \tag{5}$$

Also, each loop contains only one degree of freedom. Thus, it results:

$$M = a_1 + a_{23} \tag{6}$$

Integer solutions of the equations:

$$\begin{aligned} M - 6 + a_2 &= 0 \\ a_1 - M + a_2 &= 0 \end{aligned} \tag{7}$$

gives all variants of parallel mechanisms with assumed hypothesis.

The system of equations (7) has many solutions. Also, if other parameters are taken into consideration (the order of the joints in the loop, the geometrical parameters of the loops etc) the topology problem becomes very complex. The relation (7) defines the topology of parallel robots in a modular manner. Table 1 presents variants of PARTENR robots, solutions of (7), with a3 =0 and $6 \geq M \leq 3$.

TABLE I.
SOLUTIONS FOR PARALLEL ROBOTS

No	M	a1	a2	Remarks
1	6	6	0	Stewart Platform
2	5	4	1	
3	4	2	2	
4	3	0	3	

Also, figure 2 presents kinematic loops of those variants (figure 2a PARTENR robots with 6 dof, figure 2b PARTENR robots

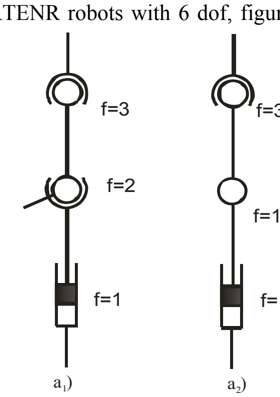


Fig. 1 Loops of parallel robots

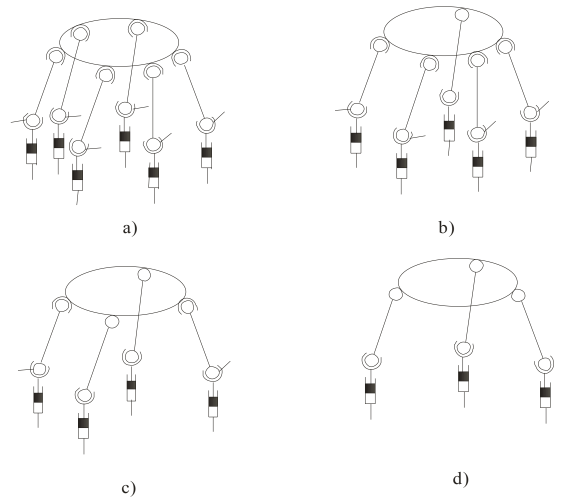


Fig. 2 Parallel robots, variants of PARTNER Robotic System.

with 5 dof, figure 2c PARTENR robots with 4 dof, figure 2d PARTENR robots with 3 dof).

C. Kinematics

General algorithms used to solve direct kinematics in the case of parallel mechanisms consider that for each independent loop of the mechanism one vector equation can be writing [5, 6]. Thus, a nonlinear system of scalar equations is obtained. Usually, this system of equations can be solved only with numerical methods and for that an accurate initial value of the solution it is required. Of course, this initial value of the solution is strongly related to the geometric parameters of the mechanism. When the geometric parameters of the mechanism are changed also the initial solution must be changed. According to that, the kinematics of the parallel mechanism will be developed in a modular manner, based on kinematics of the legs which connect the platforms and in order to ensure an analytical value for the initial solution. Each leg is in fact the right (or left) side of one independent closed loop and can be described by two coordinate systems: one attached to the frame and the other one attached to the mobile platform (Fig. 3).

The relationship between these coordinate systems is given by:

$$H_{iml} = \prod A_{il}(q_{il}) \tag{8}$$

for the left part of the independent loop and :

$$H_{imr} = \prod A_{ir}(q_{ir}) \tag{9}$$

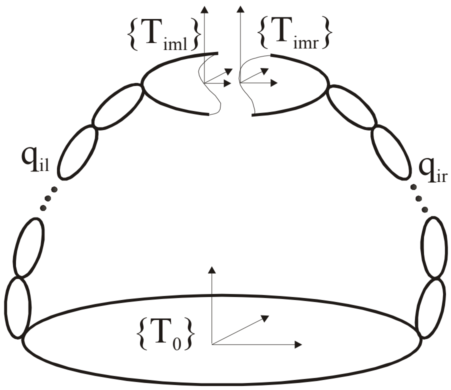


Fig. 3. "Cut body" method

for the right part. $\mathbf{H}_{iml}$, $\mathbf{H}_{imr}$ are absolute transformation matrices and $\mathbf{A}_{il}(q_{il})$, $\mathbf{A}_{ir}(q_{ir})$ are relative transformation matrices.

For an independent loop it results:

$$H_{iml} = H_{imr} \tag{10}$$

Matrix equation (10) leads to six independent scalar equations. For whole parallel mechanism, a nonlinear system of equations (with 6n independent scalar equations, where n is the number of independent loops) will be obtained. This system of equations can be solved only with numerical methods. Generally, the legs of the parallel component have the same topology. It results that the relative transformation matrices for the left and right part of each loop are formal similar. Therefore, for each topology of the legs, a formal mathematical entity (named LMM - Leg Mathematical Model) was developed. Similarly a modular kineto-static model was also developed. This mathematical model leads to non-linear system of equations. Classic algorithms of numerical methods, e.g. Newton-Raphson, can be used in order to solve this system of equations.

Usually a virtual model must be designed in order to ensure a friendly way to cooperate with the customer. Related to the virtual parallel mechanisms and in order to ensure this property, the virtual model of LMM must include an automatic way to find an initial solution for the nonlinear system of equations. Without lose the generality of the problem, a leg with PUS topology is considered (Fig.4a). An analytical solution of the initial values of the angular parameters of the joints of the leg means that a solution of the inverse geometric model for the initial position must be determined. This solution is also the initial solution for the nonlinear system of equations for the whole mechanism. Thus, for the leg from figure 4a, it results:

$$
\begin{aligned}
&H = A_1 \ldots A_6 \\
&A_1^{-1} H = A_2 \ldots A_6 \\
&A_2^{-1} A_1^{-1} H = A_3 \ldots A_6 \\
&A_3^{-1} A_2^{-1} A_1^{-1} H = A_4 \ldots A_6 \\
&A_4^{-1} A_3^{-1} A_2^{-1} A_1^{-1} H = A_5 \cdot A_6 \\
&A_5^{-1} A_4^{-1} A_3^{-1} A_2^{-1} A_1^{-1} H = A_6
\end{aligned} \tag{11}
$$

where H is the absolute transformation matrix, which describe the absolute position and orientation of the mobile platform (known for the initial position of the mechanism), Ai (i=1,6) is the relative transformation matrix. The elements of the Ai are functions of the joint coordinate (qi for the prismatic joint and ∀ij for all other joints of the leg). Using relations (11) a set of initial values for the parameters which describe the leg from figure 4a can be found.

D. Basic Description of PARTNER Robots

PARTNER Robots are parallel robots designed into modular manner, in order to ensure full reconfigurability according to customer wishes. This type of robots can be used for industrial environment but not only. Applications such as "second hand" into surgical intervention and / or active compliance devices may be common tasks for the PARTNER robots.

Modularity property of PARTNER is given mainly by the mathematic development (theoretic topology, kinematics, dynamics, control etc), in the way that this was made into modular manner. Thus, for kinematics "cut-body" method was used in order to describe closed loops kinematics. Dynamics of PARTNER robots was developed also in modular manner, using d'Alambert principle. It results that the entire topology, kinematics and dynamics of the robot becomes a problem of combination of the "mathematical" modules according to specific rules of the multi body systems. Of course, such type of complex systems must be developed into virtual models as useful instrument of optimization and / or avoiding mistakes. The modular mathematics of PARTNER was in MOBILE software package [5] implemented and few virtual models for few independent loops were obtained. Reconfigurability of the PARTNER robots results from an appropriate design of the mechanical architecture.

III. Laser Welding

A laser is a device that emits light (electromagnetic radiation) through a process called stimulated emission. The term "laser" is an acronym for Light Amplification by Stimulated Emission of Radiation. In manufacturing, lasers are used for cutting, bending, and welding metal and other materials, and for "marking"— producing visible patterns such as letters by changing the properties of a material or by inscribing its surface.

A laser consists of a gain medium inside a highly reflective optical cavity, as well as a means to supply energy to the gain medium. The gain medium is a material with properties that allow it to amplify light by stimulated emission.

Types of lasers and operating principles:

- Gas lasers
  - Chemical lasers
  - Excimer lasers
- Solid-state lasers
  - Fiber-hosted lasers
  - Photonic crystal lasers
  - Semiconductor lasers
- Dye lasers
- Free electron lasers
- Exotic laser media

Modes of operation:

- Continuous wave operation
- Pulsed operation
  - Q-switching
  - Mode locking
  - Pulsed pumping

Haas Laser HL 124P is a pulsed, Nd: YAG solid -state operated laser. Pulsed lasers provide a high power burst of energy for a short period and are very effective in some laser cutting and welding processes.

Pulsed laser operation consists into pumping the laser material with a source that is itself pulsed, either through electronic charging in the case of flashlamps, or another laser which is already pulsed.

Nd:YAG (neodymium-dopedyttrium aluminium garnet; $Nd:Y_3Al_5O_{12}$) is a crystal that is used as a lasing medium for solid-state lasers.

Characteristics for Haas Laser HL 124P:

- Laser wave length 1064 nm
- Maximum average power 120 W
- Minimum pulse power 300 W
- Maximum pulse power 5000 W
- Pulse time for minimum power 0.3..10 ms
- Pulse time for maximum power 0.3..20 ms
- Maximum pulse energy 50 J
- Maximum pulse frequency 300 Hz

Main parts of Haas Laser HL 124P:

- Laser
- Optical component
- Cooling unit
- Fiber optic cable
- Processing optic unit

IV. CONTOL SYSTEM

Centralized control of the welding cell is done using a programmable logic controller (PLC). A PLC is micro-processor based architecture used for automation of industrial processes. PLCs are used in many different industries and machines they are designed for multiple inputs and output arrangements, extended temperature ranges, immunity to electrical noise, and resistance to vibration and impact. A PLC is an example of a real time system since output results must be produced in response to input conditions within a specified time, otherwise unintended operation will result.

In contrast to PCs, which are designed to do many jobs in parallel, programmable controllers are intended to perform only one task. Compared to PCs they have many advantages:

- Hardware longevity
- System stability
- Almost instant boot-up

PLCs are available from 8 I/O to 80,000 I/O so a price point can be found for the most cost effective control. Using PCs for control task requires high processing capabilities. For real-time processes also real-time operating systems are needed (like QNX).

PLCs present also some disadvantages:

- Limited processing capability
- No true HMI capabilities
- Low level programming languages

The PLC chosen to control the PARTNER robot is a Schneider Telemecanique Twido PLC with the following specifications:

- 20 I/O ports on base controller
- Expandable with 8 modules
- Modbus, AS-I and CANOpen fieldbuses
- 2000 memory words (expandable)
- Max 0,14us instruction execution time

Programming of the controller is done in TwidoSuite, according to IEC61131-3 standard. IEC 61131-3 currently defines five programming languages for programmable control systems: FBD (Function block diagram), LD (Ladder diagram), ST (Structured text, similar to the Pascal programming language), IL (Instruction list, similar to assembly language) and SFC (Sequential function chart). For the Twido controller only LD and IL is available.

Motion of the PARTNER robot is ensuredd linear servo motors. A servo motor is a closed-loop control system. Closed loop control is based on feedback from the controlled process. If position is being controlled, a measure of the current position is provided back to the controller, allowing it to adjust its commands as the system responds. The servo system chosen to actuate the PARTNER robot is Wittenstein's Ternary Intelligent Linear Servo system. It's main characteristics are:

- Force up to 600 N
- Speed up to 225 mm/sec
- Repeatability +/- 0,045mm

These servo systems can be controlled via CanOpen fieldbus. On the same CanOpen fieldbus there are also two Schneider Altivar 31 frequency converters, which control the two three-phase motors driving the conveyor belts. This is also a closed loop control utilizing flux vector control technology.

Main characteristics of the Altivar 31 frequency converter:

- Motor and drive protection
- Linear, S, U and customized acceleration and deceleration ramps
- Jog speed option
- Brake sequence
- Automatic catching a spinning load with speed detection and automatic restart
- Fault configuration and stop type configuration
- Saving the configuration in the drive
- CANOpen and Modbus fieldbuses
- Flux vector control

Functioning is presented in figure 4. A PC connected to the PLC via RS232 (or ethernet) using Modbus-RTU (or Modbus-TCP) protocol is the main part of the human machine interface. A program specially designed and created for this application is running on the computer. It has the role of a supervisory control and data acquisition (SCADA) system. SCADA systems have 3 main parts:

Communication part, for communication with the PLC. In this case Modbus-RTU protocol is used based on the RS232 interface.

User interface, also called Human Machine Interface (HMI) is basically the part that the user sees. This is used for input of the welding point coordinates, delay times, conveyor speeds. Data are then transmitted to the PLC to ensure correct functioning even with the PC disconnected due to software or hardware problems.

Database server, for event logging, very important in automotive applications.

The role of the central PLC is to give reference to the closed-loop controls for the positioning motors and to the conveyor belt drives. It starts and stops the laser source, and verifies the laser source' response. It also collects data from process sensors, and commands any additional subsystems (warning lights, access and safety sensors and actuators).

The laser source is configured via its operator panel. Here user has the possibility to set the welding parameters. The pulse frequency and the pulse energy can be set.

The welding lens is mounted on the PARTNER robots mobile platform. Its position and orientation for each welding point is given by the user via the HMI and references for the linear servo motors are calculated for each welding point.

The conveyor belts speed is also given by the user. Giving separate speeds for the conveyors is possible. Constant speeds are ensured by flux vector control frequency converters.

V. CONCLUSIONS

In fig.5 a picture of the Partner robot, the laser source and the conveyor belt is presented.

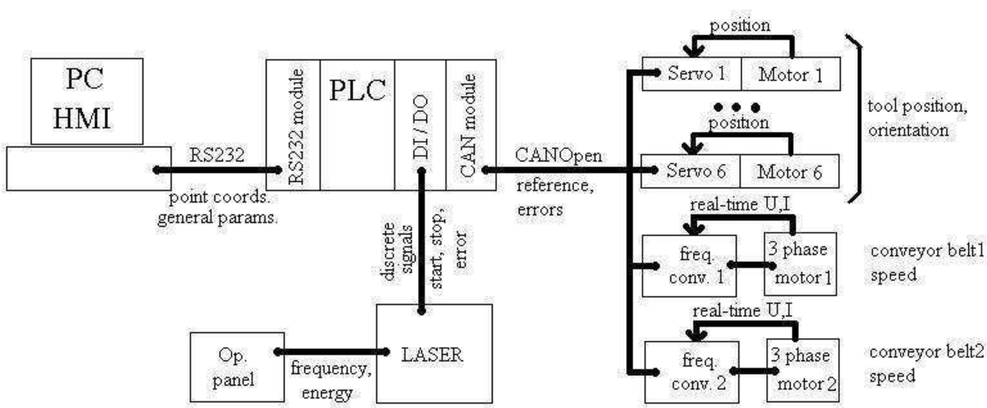


Fig. 4 Cell Overview

Fig. 5 Partner robot, laser source, conveyor belt

In our opinion the optimal manufacturing systems are the reconfigurable manufacturing systems. The laser welding cell presented in this paper fulfills the reconfigurable manufacturing system specifications:

Modularity: The mechanical structure of the cell is truly modular, the reconfigurabilty of the Partner robot is described in the next paragraph. The control system is also modular, based on a modular PLC and the CANOpen fieldbus.

Integrability: The combination of the PC and PLC in the control system ensures easy integration into any environment, with almost any industrial and non-industrial communication protocols available in a modular manner.

Convertibility: The presented laser welding cell can be converted to other type of manufacturing cell (drilling cell, screw assembly cell, milling cell) with changing of specific parts of the software and the tool.

Diagnosability: A good human machine interface with adequate PLC software is responsible for error management. Error and alarm conditions show up on the user interface and are easily interpretable.

Customization: Satisfaction of individual customer needs is possible within certain limits.

REFRENCES

1. Angeles, J., "Fundamentals of Robotic, Mechanical System", *Springer – Verlag, 1997.*
2. Borges, J." PC vs. PLC for Machine and Process Control", *Real-Time Magazine, vol.4 1997, p71-72*
3. Boterenbrood, H.: "CANopen high-level protocol for CAN-bus", *Version 3.0, NIKHEF internal documentation, 2000*
4. Brisan C., Handra-Luca, V., "Contributions to the Kinematic Structural Synthesis of the Parallel Robots", *International Conference on PKM, Ann Arbor, Michigan, USA, 2000, pp.71-78.*
5. Carretero, J.A., s.a., "Kinematic Analysis And Optimization of a New Three-Degree-of-Freedom Spatial Parallel Manipulator", *ASME Journal of Mechanical Design, Vol. 122, No. 1, 2000, pp. 17-24.*
6. Hiller, M., "Multiloop Kinematic Chains and Dynamics of Multiloop Systems, in Kinematics and Dynamics of Multi-Body Systems" (ed. By J.Angeles and A. Kecskemethy), *CISM Courses and Lecture, nr. 360, Springer Verlag, 1995.*
7. Kecskemethy, A.,"MOBILE -User's Guide and Reference Manual" Gerhard-Mervator-Universität, GH Duisburg. 1994.
8. Mackay, S. G., Park, G. M.: "Data Communications & Networking", IDC technologies, 2002
9. Mehrabi, M.G., Ulsoy, A.G., Koren, Y., "Reconfigurable Manufacturing Systems: Key to Future Manufacturing"
10. Parenti, V., Castelli, A, "Classification and kinematic modelling of fully-parrallel manipulators", *A Review Parallel Kinematic Machines: Theoretical Aspects and Industrial Requirements. 1999, pp.51-68.*
11. Pateli, A.J.,.Ehman, K.F, "Volumetric Error Analysis of a Stewart Platform Based Machine Tool", *Annals of the CIRP, Vol 46,1997, pp287-290*
12. Pritschow, G., Tran, T.L., "Parallel Kinematics and PC-based Control System for Machine Tools", *Proceedings of 37th IEEE Conference on Decision and Control, 16./18. December 1998, Tampa/Florida, USA, pp.2605-2610.*
13. Pritshow, G.,.Wurst, K.-H, "Systematic design of hexapods and other parallel link systems", *Annals of the CIRP Vol 46, 1997, pp.291-295.*
14. Pritshow, G.,.Wurst, K.-H, "Modular Robots for flexible assembly", *Proceedings of the 28th CIRP International Seminar on Manufacturing Systems "Advances in Manufacturing Technology - Focus on Assembly Systems, May 15-17, 1996..*
15. Song J. s.a., "Error Modeling and Compensation for Parallel Kinematic Machines", *Parallel Kinematic Machines: Theoretical Aspects and Industrial Requirements. 1999, pp.171-187.*
16. Ternary Intelligent Rotary / Linear Servo System, "Operating and Command instructions*", Doc No 1091-320780 Ver. 1.04, 2004*
17. Twido Main Catalog, 2006
18. Wurst, K.-H., LINAPOD, "Machine Tools as Parallel Link Systems Based on a Modular Design", *Proceedings of 1st European-American Forum on Parallel Kinematic Machines,1998, Mariland*

Generalized Net Model of E-learning System with Privacy Protection Module

Daniela A. Orozova, Veselina G. Jecheva

***Abstract*— The present paper introduces a generalized net model of e-learning system, which can satisfy some specific security requirements. The purpose is to analyze the major types of e-learning users and their roles into the distance learning process, as well as their interactions through the system. In addition, some security requirements to the e-learning system are considered and a security monitor is incorporated into the proposed model.**

***Index Terms*—e-learning, generalized net, privacy, Moodle.**

I. Introduction

E-learning is a network-enabled type of education, where the transfer of skills and knowledge is accomplished via the Internet, LAN/WAN, audio or video tape, satellite TV, and CD-ROM. All universities include e-learning techniques into their traditional and distance education. In many cases the term e-learning denotes distance learning through internet or web based training. E-learning is learner-centric, personalized, and interactive. Usually the e-learning systems are available anytime, at any place, at the convenience of the student.

Usually e-learning systems have been considered as Course Management Systems (CMS), Learning Management Systems (LMS) or Virtual Learning Environments (VLE). Their major purpose is to be a common environment, focused on administrative aspects of learning and on content delivery.

The basic components of an e-learning process are [9]:

- Students - people, who use e-learning system to achieve learning goal;
- Teachers - people, who use e-learning system in order to monitor, guide, help and assist students when trying to achieve learning goal;
- E-content – presents learning goal and should describe the way, how that goal can be achieved;
- Technology - e-learning environment should be accessed using a web browser over the internet or intranet and allows users to access and deliver training content.

Manuscript received October 15, 2008. This work was supported by Burgas Free University under Project 10/2008 and Faculty of Computer Science and Engineering.

D. A. Orozova is with Burgas Free University, Burgas, Bulgaria (phone: +359-56-900-531; e-mail: orozova@ bfu.bg).

V. G. Jecheva is with Burgas Free University, Burgas, Bulgaria (phone: +359-56-900-487; e-mail: vessi@ bfu.bg).

The e-learning service offers collaborative communication environment the students and faculty can communicate through [12]. This communication could be direct (messages, forums, chats) or indirect (course resources, tests, interactive lessons, wikis, assessments, etc.). Each user indicates with his username and password. Behind each virtual name lies the user profile, containing the user status (role) in the system, and personally identifiable information, such as e-mail address, real name, photo, homepage URL, home or work address or telephone number. The user profile also contains some demographic information about the user, which is not unique to user, such as ZIP code, age, gender, preferences, interests and hobbies. In addition, most e-learning systems automatically collect information about user hardware and software, like IP address, browser type, access times and referring Web site addresses. The user personalization is critical, since the learning content has to change dynamically, according to the student personal preferences or needs. The reason is that each student has different learning style preferences, and various past learning experiences [13].

These features of e-learning systems enforce some security and privacy requirements. Like all public network systems, e-learning systems are exposed to attacks. Students may access the system using many different PCs or mobile devices and from various locations [16]. The security module has to be responsible for regular updating and monitoring of the e-learning system, as well as following the requirements of security measures for access control, information confidentiality and privacy [19]. The major privacy elements of e-learning system are anonymity: no user has to reveal the private information of another one without permission; pseudonyms: each user works with their virtual name; and authenticated usage: user secure authentication is required [17]. So the system has to provide secure information flows, i.e. each resource has to be accessed only by users, who have granted access to it.

The purpose of the present paper is to examine the information flows in sample e-learning system and create a generalized net model, which represents the system as a communication environment. The created model takes into consideration privacy and access control requirements to the system.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_57, © Springer Science+Business Media B.V. 2010

II. THE CASE STUDY

As an exemplary e-learning system we use Moodle [18] with a purpose to supplement traditional university learning of both full-time and part-time students. It is a free, open source course management system (CMS), that is developed for delivery and using the best educational practices. Moodle promotes social constructionist pedagogy principles, which include collaboration, activity-based learning, critical reflection, etc. This solution is designed in a modular way and places at user disposal flexible tools to add or remove functionality at many levels.

Moodle contains some predefined roles, which are built by default, and allows various roles to be created. The major roles, that we use, are “Administrator”, "Teacher", "Student" and “Guest".

- Administrators (content publishers) can change all system preferences and permissions, create and delete courses and grant access and assign roles to other users. They have physical access to the resources of the system and can alter the parameters of the system servers and databases. In our case the user-side registration in the system is disabled, so the administrators add and remove users from the database.
- Teachers (content creators) can create course content, which includes various resources, presenting learning materials, homeworks, instructions, additional readings, URLs; add and remove course agenda, create tests, publish news and announcements, related to the course and create and participate into interactive forms of contact with students, like forums and chats. Teachers also manage student enrolments, track students activity and progress in the system. They are responsible not only for the practical secure realization of the e-learning site but also for putting together all e-learning objects and the use of additional elements.
- Students (end users) can access all available resources, solve tests and quizzes, fulfill assignments and participate into forums and chats, if any.
- Guests have minimum rights and their role can be disabled. Usually they can view few resources, which are given open access to all users.

In addition, the system administrators can create custom roles with their specific rights.

In this collaborative learning environment teachers’ and students’ roles are interdependent. The students are not passive consumers of information, but need to become self-reliant, active searchers for relevant information [5]. They can collaborate with other students and the teacher into some project-based activities, as well as cooperate in team work. Depending on the course concept, students can even teach other students, using social and active learning principles.

III. THE GENERALIZED NET MODEL

A. Generalized Nets

Petri nets, introduced by Petri [1] in the 60s, are very popular mathematical tools for distributed systems modeling. Since this time they have been modified and developed many times. The Colored Petri nets are among their most popular extensions [14]. Generalized nets (GN, see [2]) include as particular case Petri nets and all their modifications. They are introduced in [3] as general modeling tool, intended for abstract objects (neural networks, genetic algorithms, expert systems, abstract systems, etc.). In [7, 10] is introduced a GN model, which is focused on the assessments of the students’ knowledge, based on the fuzzy sets evaluations.

The major advantages of GN over the other similar models (Petri nets and their modifications) are the following:

- The GN tokens are objects, which enter the net with initial characteristics. These characteristics can change their values during the tokens transitions in the net.
- Each particular transition has an associated matrix, which elements are predicates. The values of these predicates determine the directions of tokens movement from the starting to the end positions.
- The time of each GN can be measured using a preliminarily defined time scale. When two or more GN work in parallel, their combined time can be calculated jointly.

The first basic difference between GNs and the ordinary Petri nets is the "place - transition" relation. Here, transitions are objects of a more complex nature. Formally, every GN-transition is described by a seven-tuple, but here we shall discuss its reduced form:

Z = < L', L", r >,

where:

a) L' and L" are finite, non-empty sets of places (the transition's input and output places, respectively);

b) r is the transition's condition determining the tokens which will transfer from the transition's inputs to its outputs; it has the form of an indexed matrix (see Ref. [2]):

$$r = \begin{array}{c|ccccc} & l_1'' & \ldots & l_j'' & \ldots & l_n'' \\ \hline l_1' & & & & & \\ \ldots & & & m_{i,j} & & \\ l_i' & & & & & \\ \ldots & & & & & \\ l_m' & & & & & \end{array}$$

where $r_{i,j}$ $(1 \le i \le m, 1 \le j \le n)$ are predicates and (i,j) denotes the element which corresponds to the i-th input and j-th output places; these elements are predicates and when the truth value of the (i,j)-th element is true, the token from i-th input place can be transferred to j-th output place; otherwise, this action is not possible.

The ordered four-tuple (also for the reducing case):

E= < A, K, X, *>

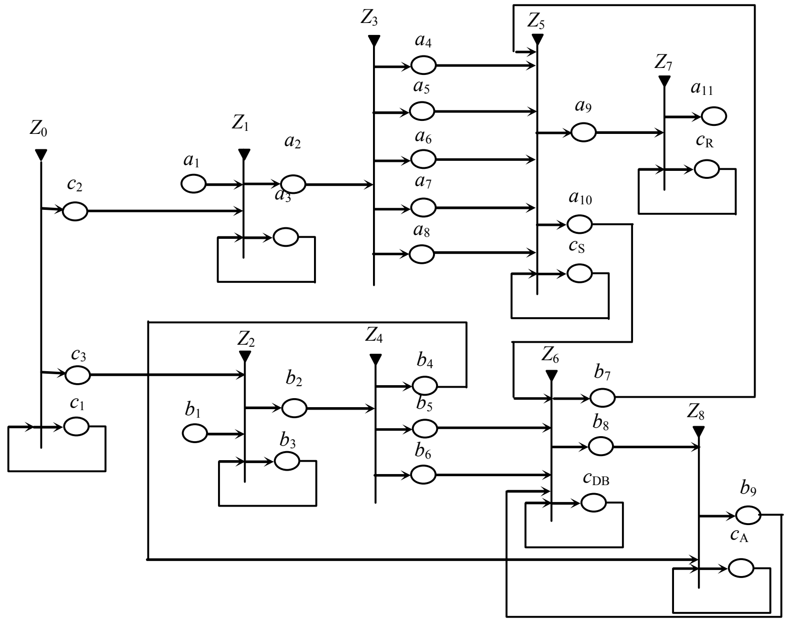


Fig. 1. GN-model of e-learning system with privacy module and access control.

is called a reduced Generalized Net, if:

a) A is a set of transitions;

b) K is the set of the GN's tokens;

c) X is the set of all initial characteristics the tokens can receive on entering the net;

d) * is a characteristic function, which gives new characteristic to every token when it makes a transfer from an input to an output place of a given transition.

B. The E-learning System Model

Here we construct a Generalized Net, shown on Fig. 1, which represents the communication process in the developed system. This model allows to track the changes in the objects parameters and to monitor the system state.

The generalized network design problem has several applications to telecommunication and location problems [11], railway optimization [6], etc. Ref. [2] represents an application of GN models for sending and receiving electronic data in World Wide Web. In Ref. [4] is proposed a method of teaching quality estimation at universities and GN-model of server is described.

The constructed GN-model has the following tokens:

- α-tokens, representing users with non-administrative rights: students, teachers and guests and their actions;
- β-tokens, representing administrators;
- δ-tokens, representing the information in the databases, containing learning and additional resources of the system;
- γ-tokens, representing security module of the system.

For the purpose of simplicity we will use the denotations α, β- tokens instead of α_i- and β_j-tokens, where *i, j* are the numbers of the respective tokens. In place c_1 there is one γ-token with initial and current characteristic:

$$x_0^{\gamma} = \text{“Privacy module and access control”}.$$

The γ-token generates two new tokens, that enter in places c_2 and c_3 and they not obtain new characteristics.

Initially α- and β-tokens are in places a_3 and b_3 respectively with initial characteristics:

$$x_0^{\alpha} = \text{“(student/guest/teacher), password, name, status, access rights”},$$

$$x_0^{\beta} = \text{“password, name, status (course administrator or global administrator)”}.$$

All α-tokens have equal priorities and all β-tokens have equal priorities but the priority of the β-tokens is higher than the priority of α-tokens.

New users and administrators enter the net through places a_1 and b_1, respectively.

During the functioning of the net the following tokens with current characteristics are always present: δ_S-token in place c_S with characteristic “DB with messages”; δ_R-token in place c_R with characteristic “Trash”; δ_{DB}-token in place c_{DB} with characteristic “DB”; δ_A-token in place c_A with characteristic “archive”.

The transitions are of the described below type:

$$Z_0 = \langle \{c_1\}, \{c_1, c_2, c_3\}, \begin{array}{c|ccc} & c_1 & c_2 & c_3 \\ \hline c_1 & true & true & true \end{array} \rangle,$$

$$Z_1 = \langle \{a_1, a_3, c_2\}, \{a_2, a_3\}, \begin{array}{c|cc} & a_2 & a_3 \\ \hline a_1 & false & true \\ a_3 & W^a_{3,2} & W^a_{3,3} \\ c_2 & false & true \end{array} \rangle,$$

$W^a_{3,2}$ = "The user will take actions in the e-learning system.",

$W^a_{3,3} = \neg W^a_{3,2}$, where $\neg$P is the negation of the predicate P.

α-tokens do not receive new characteristics in place a_3 and receive the characteristic "list of all allowed actions of the user in the system" in place a_2.

$$Z_2 = \langle \{b_1, b_3, c_3\}, \{b_2, b_3\}, \begin{array}{c|cc} & b_2 & b_3 \\ \hline b_1 & false & true \\ b_3 & W^b_{3,2} & W^b_{3,3} \\ c_3 & false & true \end{array} \rangle,$$

$W^b_{3,2}$ = "The administrator will initiate actions in the e-learning system",

$W^b_{3,3} = \neg W^b_{3,2}$.

β-tokens do not receive new characteristics in place b_3 and receive characteristic "list of possible administrator actions in the system" in place b_2.

$$Z_3 = \langle \{a_2\}, \{a_4, a_5, a_6, a_7, a_8\}, \begin{array}{c|ccccc} & a_4 & a_5 & a_6 & a_7 & a_8 \\ \hline a_2 & W^a_{2,4} & W^a_{2,5} & W^a_{2,6} & W^a_{2,7} & W^a_{2,8} \end{array} \rangle,$$

$W^a_{2,4}$ = "The action Sending a message or resource to another user in the system has been chosen",

$W^a_{2,5}$ = "The action Review online all the received messages or resources has been chosen"

$W^a_{2,6}$ = "The action Deletion of received messages or resources has been chosen",

$W^a_{2,7}$ = "The action Review the messages or resource in Trash has been chosen",

$W^a_{2,8}$ = "The action Filtering messages has been chosen".

α-tokens, entering in places a_4, a_5, a_6, a_7 and a_8 receive the respective characteristics:

- "user, action Sending a message or resource to another user in the system",
- "user, action Review online all the received messages or resource",
- "user, action Deletion of received messages or resource",
- "user, action Review the messages in Trash or resource",
- "user, action Filtering messages".

$$Z_4 = \langle \{b_2\}, \{b_4, b_5, b_6\}, \begin{array}{c|ccc} & b_4 & b_5 & b_6 \\ \hline b_2 & W^b_{2,4} & W^b_{2,5} & W^b_{2,6} \end{array} \rangle,$$

$W^b_{2,4}$ = "Modifications in the Databases containing learning and additional resources of the system are required ",

$W^b_{2,5}$ = "Archivation of the Database containing learning and additional resources of the system is required",

$W^b_{2,6}$ = "The retrieval of information from the archive is required ",

β-tokens, entering in places b_4, b_5 and b_6 receive the respective characteristics: "administrator, action Modifications in the Databases", "administrator, action Archivation of the Database", "administrator, action retrieval of information from the archive".

$$Z_5 = \langle \{a_4, a_5, a_6, a_7, a_8, b_8, c_S\}, \{a_9, a_{10}, c_1\}, \begin{array}{c|ccc} & a_9 & a_{10} & c_S \\ \hline a_4 & false & false & true \\ a_5 & false & false & true \\ a_6 & false & false & true \\ a_7 & false & false & true \\ a_8 & false & false & true \\ b_8 & false & W^c_{S,10} & true \\ c_S & W^c_{S,9} & false & true \end{array} \rangle,$$

$W^c_{S,9}$ = "Deleting messages or resources from Trash is required";

$W^c_{S,10}$ = "Access to DB containing learning and additional resources of the system is required".

α-tokens, entering in places a_9 and a_{10} receive characteristics respectively: "user, action Deletion from Trash"; "user, access request".

$$Z_6 = \langle \{a_{10}, b_5, b_6, b_9, c_{DB}, b_{10}\}, \{b_7, b_8, c_{DB}\}, \begin{array}{c|ccc} & b_7 & b_8 & c_{DB} \\ \hline a_{10} & false & false & true \\ b_5 & W^b_{5,7} & false & W^b_{DB} \\ b_6 & false & false & true \\ b_9 & false & false & true \\ c_{DB} & false & W^b_{DB} & true \end{array} \rangle,$$

$W^b_{5,7}$ = "Modifications in the message database are needed",

W^b_{DB} = "Modifications in the database, containing learning and additional resources of the system are needed".

β-tokens, entering in places b_7 and b_8 receive the following characteristics respectively: "administrator, action Modifications in message database", "administrator, action Archivation of database".

$$Z_7 = \langle \{a_9, c_R\}, \{a_{10}, c_R\}, \begin{array}{c|cc} & a_{10} & c_R \\ \hline a_9 & false & true \\ c_R & W^c_{R,10} & true \end{array} \rangle,$$

$W^c_{R,10}$ = "Deletion of the information from Trash is required".

α-tokens entering in place a_{10} receive characteristic "user, deleted information".

$$Z_8 = \langle \{b_4, b_8, c_A\}, \{b_9, c_A\}, \begin{array}{c|cc} & b_9 & c_A \\ \hline b_4 & false & true \\ b_8 & false & true \\ c_A & W^c_{A,9} & true \end{array} \rangle,$$

$W^c_{A,9}$ = "Information retrieved from the archive".

IV. Conclusion

The major purposes of combining e-learning techniques to the traditional education are: to get better teaching, more motivated students, and more well structured and interactive courses. This paper represents a generalized net model, which describes the information flows in e-learning system, where some privacy and access control requirements are observed. This generalized net model could be used for future investigations in order to add more participant roles and various security requirements. Using this model, e-learning systems can be better and effectively implemented by vendors.

The advantages of using the GN model include the following features:

- allow real-time management of the modeled system;
- allow optimization and further elaboration of the real processes.

In the presented model the transitions does not contain time components, since they can be defined as characteristics of some tokens. In some particular cases, if necessary, the time components can be added for each transition.

REFERENCES

[1] Котов В.Е., Сети Петри, "Наука" Москва, 1984.
[2] Atanassov, K., Generalized Nets, World Scientific, Singapore, 1991.
[3] Atanassov, K., H. Aladjov. Generalized Nets in Artificial Intelligence. Vol.2: Generalized Nets and Machine Learning. “Prof. M.Drinov” Academic Publishing House, Sofia, 2001.
[4] Atanassov, K., S. Sotirov, V. Kodogiannis, Intuitionistic fuzzy estimations of the Wi-Fi connections, First Int. Workshop on IFSs, GNs, KE, London, 6-7 Sept. 2006, 75-80.
[5] Esposito F., O. Licchelli, G. Semeraro, Discovering Student Models in e-Learning Systems, Journal of Universal Knowledge Management, Vol.10, Issue 1, pp. 46-57.
[6] Hadjicharalambous G., P. Pop, E. Pyrga, G, Tsaggouris, C. Zaroliagis, The Railway Traveling Salesman Problem, Algorithmic Methods for Railway Optimization, Lecture Notes in Computer Science, Vol.4359/2007, pp. 264-275.
[7] Hristova M., E. Sotirova, Generalized net model of the multifactor method of teaching quality estimation at universities, In Fourth International IEEE Conference Intelligent Systems, 2008, 16-20 – 16-24.
[8] Hristova M., E. Sotirova, Multifactor method of teaching quality estimation at universities with Intuitionistic Fuzzy Evaluation, Eleventh Int. Conf. on IFSs, Sofia, NIFS Vol. 14 (2008) 2. 80-83
[9] Kastelic M., T. Lončarič, A Model of Introducing E-learning System at Vocational College for Business Secretaries, The Journal of Issues in Informing Science and Information Technology, Volume 4, 2007, pp. 175-187.
[10] Langova-Orozova D., H. Aladjov, K. Atanasov, Generalized Net Fof Machine Learning With Current Estimations, Advanced Studies in Contemporary Mathematics, Ku-Duk Press, Vol.3, 2001, No2, pp. 61-76.
[11] Laporte G., A.A.Vaziri, C. Sriskandarajah, Some applications of the generalized traveling salesman problem, J. Oper. Res. Soc., Vol. 47, 1996, pp. 1461-1467.
[12] Lee G., Stanley Y.W. Su, Learning Object Models and an e-Learning Service Infrastructure for Virtual e-Learning Communities, Advances in Web-Based Learning – ICWL 2004, LNCS Volume 3143/2004, pp. 371-378.
[13] Raitman R.; L. Ngo; N. Augar; W. Zhou, Security in the online e-learning environment, In Proceedings of Fifth IEEE International Conference on Advanced Learning Technologies, 2005. ICALT 2005, pp. 702 – 706.
[14] Jensen, K.; Rozenberg, G., High-Level Petri Nets, Berlin, Germany: Springer-Verlag, 1991.
[15] Rosenberg M.J., E-learning: Strategies for Delivering Knowledge in the Digital Age, McGraw-Hill Professional, 2001.
[16] Warren M., W. Hutchinson, Information Security – An E-learning Problem, Advances in Web-Based Learning - ICWL 2003, pp. 21-26.
[17] Weippl E.R., A M. Tjoa, Privacy in e-learning: anonymity, pseudonyms and authenticated usage, Interactive Technology and Smart Education, Vol. 2, Issue 4, 2005, pp. 247 - 256
[18] http://www.moodle.org
[19] Zinnen, A., Goertz, M., Faatz, A., Godehardt, E. & Lokaiczyk, R., Privacy Issues when Rolling out an E-Learning Solution, In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008, Vienna, Austria, pp. 6394—6393.

Daniela A. Orozova was born in 1967 in Gorna Oriahovitza, Bulgaria. She obtained her master degree in Computer Science and Teaching from Sofia University “St. Kliment Ohridski” in 1990 and became an assistant professor in University of Sofia, Faculty of Mathematics and Informatics. She received her PhD degree in Computer Science, especially Intelligent Tutoring Systems and Databases, from Bulgarian Academy of Sciences, in 2001. She became an Associate Professor at Burgas Free University, Faculty of Computer and Technical Sciences, Bulgaria. Her research interests include Databases, Intelligent Tutoring Systems, E-learning and Distance learning, Modeling tools.

Veselina G. Jecheva was born in 1971in Burgas, Bulgaria. She obtained her master degree in Computer Science and Economics from Sofia University “St. Kliment Ohridski” in 1995. She received her PhD degree in Computer Science, especially Information Security from National Laboratory of Computer Virology, Bulgarian Academy of Sciences, in 2005. She is a lecturer at Burgas Free University, Centre of Computer and Technical Sciences, Bulgaria. Her research interests include information security, e-commerce, programming and Web systems.

Advanced Knowledge Assessment System – Functional and Conceptual Specification

Jan GENCI
Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Letna 9, 042 00 Kosice, Slovakia
e-mail: genci@tuke.sk

***Abstract*- E-learning has become an accepted technology widely used during the last decade. However, main attention was devoted to content development, its presentation and functionality within learning management systems. On the other side, there was not so high interest in the area of knowledge assessment. This contribution deals with the specification of an advanced knowledge assessment system, it includes the functional specification and the conceptual model too.**

I. INTRODUCTION

In the area of e-learning main attention was devoted to content development, (including authoring tools), content presentation and specification and implementation of functionality of learning management systems (LMS) during last ten years. These aspects were thoroughly elaborated and LMS' were released in the many commercial and open-source implementations (e.g. [1], [2]). We state, that is not the case of knowledge assessment despite the fact that there exist specific system e.g. for Java knowledge evaluation [3].

Indubitably, there exist groups of professionals who are involved in the development of testing suites for knowledge assessment. Theory and methodology regarding test development is quite well-developed [4]. Many standards regarding testing were developed and published (just to mention one of it – IMS QTI [5]). Despite these facts, many of (university) teachers do not follow recommendations provided by the pedagogical theory and/or standards.

One of the reasons why the theory and the standards are not applied in the knowledge assessment processes is the fact that no accepted methodology of development of knowledge assessment tests is implemented in the current LMS. According to Lavicky [4] knowledge test should be developed by the team of the experts. Aim of the test, target group, expected result and many other attributes of the knowledge testing should be specified before the development of knowledge test starts.

Let's take just one of the most popular LMS Moodle [6]. If we use it for knowledge test development, we reveal very early, that there are no means/tools:

- for cooperative development of test;
- to find development of which testing item is finished and which one is still under development;
- concerning versioning of the test items;
- for retirement of the items;
- to approve test items;
- to do the same thing regarding test what we can not do with test items (cooperative development, versioning, retirement etc.).

More generally speaking, no test item life-cycle support and no knowledge test life-cycle support are implemented in the current LMS available on the market.

Of course, we can find the way how to bypass the mentioned drawbacks setting some organizational measures. But it needs some extra effort carried out by a team of developers. Moreover, the defined measures can be easily violated and they cannot be check by the LMS.

Test developers use very often only the basic types of testing items. Every LMS implements a few types of items for testing purposes (true/false, one-from-many, many-from-many); in the more powerful systems question types as ordering, fill-in-the-blank, matching, short answer/essay can be found.

Developed tests are very often "flat" (without structure corresponding to the curriculum), composed as a "random" subset of questions and usually tightly coupled to basic knowledge.

At the end, LMS lack a functionality concerning statistics of tests and testing items used in the real tests (reporting functionality).

From the organizational point of view, tests are usually developed by course developers who are usually not aware the fact that tests have to be developed carefully and "tuned" afterwards. Every testing item has to be explored in details in order to reveal and remove "hidden" meaning in the testing question and presented answers.

Another problem regarding knowledge assessment consists in the "complex" testing (i.e. midterm or final knowledge assessment). Item is usually designed to prove the fact the students were able to remember information from the presented chapter. Final course assessment (testing) consists of mixture of items from all chapters. Higher levels of testing questions (which are able to prove higher-level of understanding of course materials) are not so frequent.

Authors' personal e-learning experience in the frame of the one of the largest e-learning project - CISCO Networking Academy (CNA)[7] program and also as a university teacher proved another aspect of testing – testing questions are usually very quickly revealed after their first use. In the global systems, what the CNA program is, students are very often

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_58, © Springer Science+Business Media B.V. 2010

able to find questions and answers on the Internet. "Local test" (developed in-house by teachers) are "stolen" using current edge technology (e.g. digital cameras, mobile-phones or even simple by cut&paste procedure). In these cases, results do not mirror students' real knowledge and testing is carried out only formally.

We are aware that not all type of knowledge can be assessed by computerized test. However, our experience in the area of information technology education convinced us in the possibility to fight with the above-mentioned problems and reach quite high level quality of knowledge assessment.

II. Advanced Knowledge Assessment Systems – Requirements

To address the drawbacks presented in the previous section, we briefly summarize them:

- impossibility of test items to demonstrate level of understanding of course material;
- absence of definition of curriculum structure (model) in the LMS;
- poor set of types of items implemented in the currently available LMS;
- personalization of the test.

To assist developers of tests in their work, we require the following functionality of a LMS in the area of testing:

- support of test item lifecycle;
- easy and flexible ways of generation of tests;
- test lifecycle support;
- easy inclusion of new type of testing questions;
- supporting the collaboration of developers of test;
- extensive reporting possibilities (e.g. regarding items, learning objects, whole test).

In the next chapter we will try to introduce technologies and approaches that we consider helpful in the area of knowledge assessment.

III. Advanced Knowledge Assessment Systems – Technologies

To alleviate, or even eliminate, drawbacks of current LMS in the area of knowledge assessment, the approaches described bellow we consider as helpful.

Taxonomy of the cognitive aims

Human being passes several phases of knowledge acquisition during the learning process – starting from the remembering, going through the understanding and using it in the productive activities, at the end. Learning process has to pass these phases which are usually named as cognitive aims.

There are several specifications of levels of cognitive aims. We call it taxonomies. Well known is Bloom's taxonomy [8-4], which defines six levels of knowledge and intellectual skills:

- Knowledge: recognize or recall information;
- Comprehension: demonstrate that the student has sufficient understanding to organize and arrange material mentally;
- Application: a question that asks a student to apply previously learned information to reach an answer;
- Analysis: higher order questions that require students to think critically and in depth;
- Synthesis: higher order question that asks the student to perform original and creative thinking;
- Evaluation: a higher level question that does not have a single correct answer.

Another one taxonomy, widely used in Slovakia, is Niemierko's taxonomy [9], which defines four levels:

- Remembering – similar to knowledge in Bloom's taxonomy;
- Understanding – similar to comprehension in Bloom's taxonomy;
- Specific transfer – application of acquired information according presented patterns;
- Non-specific transfer – creative application of acquired information.

Mapping of testing questions to different levels of a taxonomy gives us a possibility to assess whether the student has reached the required level, or to find out, which level of taxonomy student has reached during the learning process.

We require that every question designed in the assessment system to has be assigned to a particular level of the chosen taxonomy. Several taxonomies, even defined by the user, should be supported by the Advance Knowledge Assessment System (AKAS).

Definition of knowledge structure of curriculum

In order to better understand the cognitive aims of a curriculum, mapping of knowledge (learning objects or learning items) and their relationships is useful. In this case we get a certain knowledge-structure of the curricula. This structure can be presented as an oriented graph.

We illustrate this on simple example. One of our course – Unix/Linux system programming – presents, among other topics, two (independent) topics - file manipulation and synchronization of processes. File manipulation topic contains learning items regarding functions open(), close(), read() and write(), while synchronization topic is built around the concept of semaphore. However, special form of the open() function can be used for synchronization purposes too. Structure of the curricula for this case is presented on the Fig. 1.

We require every item to be assigned to a particular node of the graph. It gives us possibility to specify whether an item assesses a basic or a more complex knowledge. This requires definition of structure of curriculum to be supported by the Advance Knowledge Assessment System.

At the same time we require the AKAS to provide means for classical definition of the curriculum structure. This can be used to show the examinee the relevant course material in the case testing items is answered incorrectly.

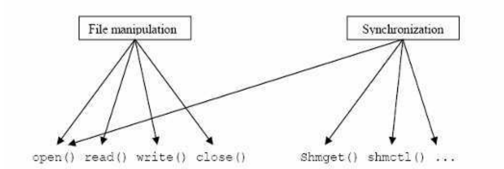


Fig. 1. Example of cognitive structure of the curriculum

RICH SET OF TYPES OF ITEMS

As mentioned above, many of the LMS provide only few types of items. It is usually: true/false, one-from-many, and many-from-many types. However, to evaluate higher level of knowledge, it will be valuable to have support of other types of "powerful" test items, i.e.:

- clickable images;
- evaluation of answer based on regular expressions;
- evaluation of answer by custom program.

Clickable images we regard as o powerful type of question in many situations. Point and click answer can be used in geography, computer architecture, computer graphics etc.

Using regular expression (RE) can, in the first plan, eliminate general typos during the item evaluation (i.e. common misspelling which is not regarded as a mistake). In the second plan, regular expressions can be more useful if we want to express large amount of possible answers in a very comprehensive way. At the end, their implementation provides possibility to use new kind of items with open answer based on regular grammars.

The possibility to incorporate evaluation program (on the testing item developer level) represents the most flexible approach. In this case, answer provided by the examinee can be quite comprehensive. We already use this approach in the Operating Systems course supported by the system [10] developed at our department. For the Operating systems course students submit source code of computer program to perform specified actions (i.e. emulated FAT file system operations [11]). For the Database Systems course students write and submit required SELECT statement in another system. Statement is executed in the real database environment and its result is compared with result of execution of SELECT statement prepared by testing item developer [12].

For all the item types (including "powerful") we require full support in the advanced assessment system. By the full support we do not mean their presence in the system only. We require appropriate design and item evaluation tools.

ANTI-CHEATING MEASURES

We believe that problem of cheating can be partially eliminated by strong personalization of the knowledge tests. It means that each student has to get her/his own version of the test. Common approach used in the current systems – changing order of items and changing order of answers in the one-from-many and many-from-many types of item – seems to be insufficient. Solution could be the higher variability of the test items – specification of much more positive and negative answers for particular question (i.e. if we specify 3 positive and 4 negative answers for one-from-four question type, overall we have 12 different combinations of possible answers for one question). Another solution could represent open answer item type evaluated by regular expression or even evaluated by specially designed computer program.

"On-line" cheating (using mobile phone or other on-line communication tools) can be hardly completely eliminated. Some technological solution can help, especially in area of restriction of on-line communication (firewalls). But they cannot be used in all situations and some of them can be limited even by legislation (e.g. local cancellation of mobile signal is not allowed in the Slovakia). Time restriction is the means, which could help to eliminate this problem. We shall require system functionality to allow to control not only overall test duration, but also to control particular questions or parts of tests.

IV. CONCEPTUAL MODEL OF ADVANCED LEARNING MANAGEMENT SYSTEM

Conceptual models try to show basic functionality of the specified system together with specification of user roles. We used UML use case diagrams to specify conceptual diagram.

During analysis of the system several roles were discovered/designed:

- administrator;
- project leader;
- developer of knowledge test;
- designer of knowledge structure of curriculum;
- developer of taxonomies;
- developer of test items.

The use case diagram of advanced knowledge assessment system is presented in the Fig. 2. We describe briefly the roles.

Administrator is responsible for maintaining the system, user account administration and system archiving. He/she grants the project leader role to users also. That is the only role granted by administrator. Any other roles are granted by project leader. Administrator can allow the project leader role to create new project, too.

Project leader can create new projects (if right is granted). He/she approves taxonomies and their mapping, grants user roles for test developer, designer of knowledge structure of curriculum, taxonomy developer and item developer.

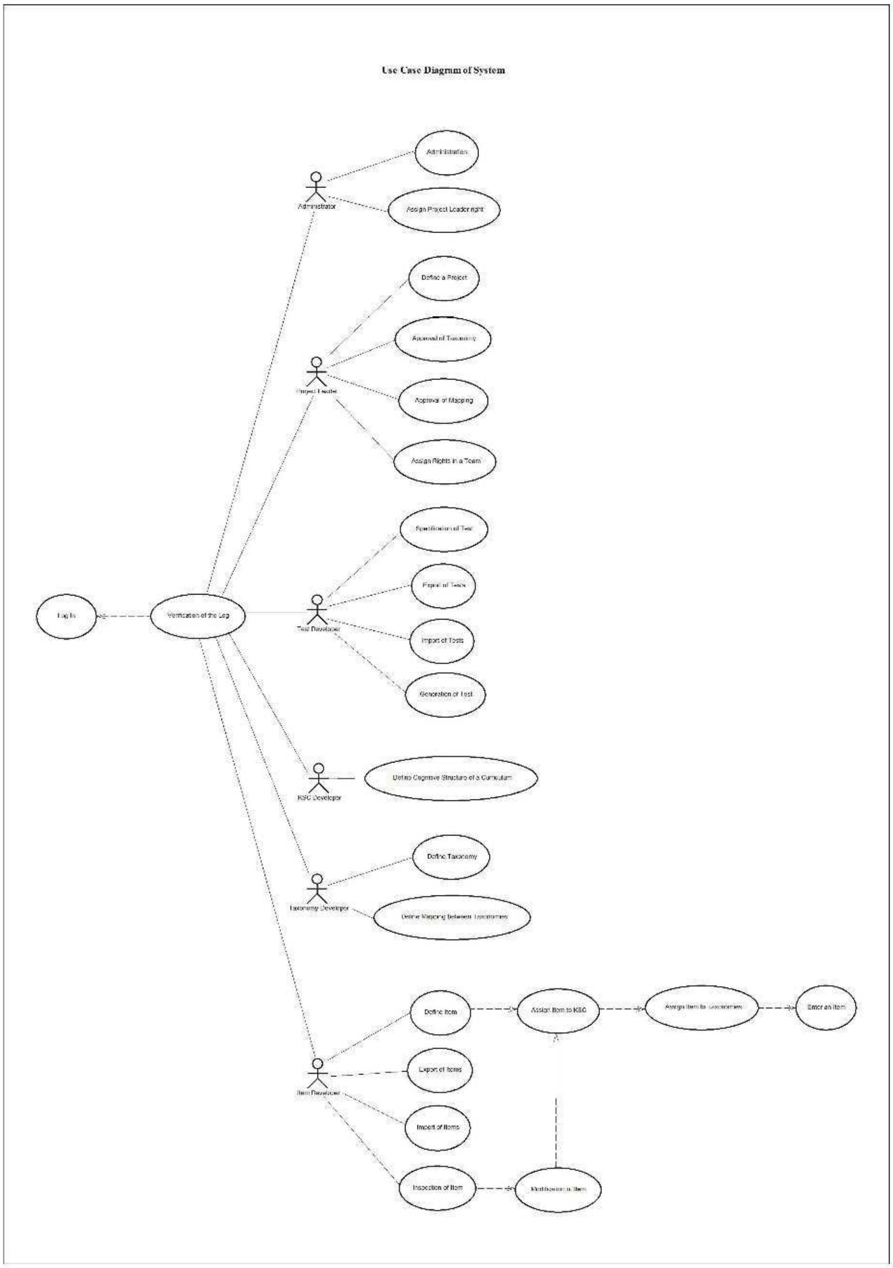


Fig. 2. Use case diagram of advanced knowledge assessment system

Test developer is responsible for knowledge test specification, its generation and setting the parameters of test.

Designer of knowledge structure of curriculum (KSC) is responsible for development of cognitive structure of curriculum. Cognitive structure of curriculum must be approved by project leader.

Taxonomy developer is responsible for development of required taxonomies and specification of mapping between taxonomies. Taxonomies and corresponding mapping has to be approved by project leader.

Developer of test items is responsible for development of testing items (question and corresponding answers). He/she is responsible for selection of correct type of question, specification of corresponding level in the supported taxonomies, "placing" item to relevant node of knowledge structure of curriculum, specifying text of question, possible answers and their parameters (correctness, rating, etc.).

The state diagram for taxonomy development is presented in the Fig. 3 and state diagram for the mapping between levels of different taxonomies is presented in the Fig. 4.

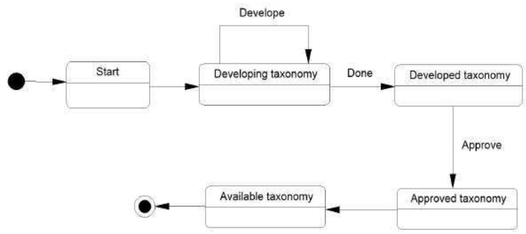


Fig. 3. Taxonomy development state diagram

V. CONCLUSION

According to the presented ideas we started design and implementation of advanced knowledge assessment system. We explored two ways:

- extension of LMS Moodle for richer question types;
- design and implementation of system from the scratch.

First approach seems to be more productive, because of enormous amount of works involved already in the LMS Moodle. However, it could be limited by functionality provided by the LMS Moodle. In the first version of the system we decided to implement functionality for collaborative development of testing item. We plan to realize the test itself in the LMS Moodle by importing testing items and test with their limited functionality.

Second approach pretends to be solution covering broader functionality presented in the paper, but will require full development of the system from the scratch.

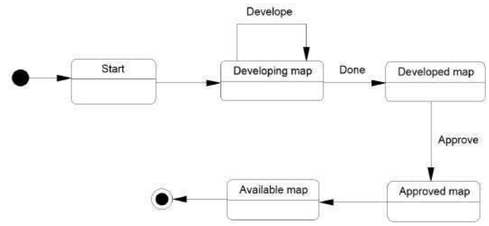


Fig. 4. State diagram - mapping between levels of different taxonomies

REFERENCES

[1] Virtual learning environment. From Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Course_management_system

[2] edutools. Decision-making tools for E-D-U community. http://www.edutools.info/

[3] Samuelis Ladislav, Szabó Csaba: Automatic the Measurement of the Complexity of Students´ Assignments, Knowledge Technologies and Applications, Networking Centres of High Quality Research on Knowledge Technologies, Košice, 31 May - 1 June, 2007, Budapest, SzTAKI, 2007, 1, pp. 116-125,

[4] Lavicky T.: Tvorba a vyuzivanie skolskych testov. (Ucebny text pre PVPZ a PV). Metodicko- Pedagogicke centrum Presov. Year not stated. (in Slovak) http://www.mcpo.sk/downloads/Publikacie/PrirodPred/PPCHE200501.pdf

[5] IMS Global Learning Consortium. Question and Test Interoperability Project Group. http://www.imsglobal.org/QTI.html

[6] Course Management System Moodle. http://www.moodle.org

[7] Cisco Networking Academy. http://cisco.netacad.net

[8] Taxonomy of Educational Objectives: The Classification of Educational Goals; pp. 201-207; B. S. Bloom (Ed.) Susan Fauer Company, Inc. 1956.

[9] Niemierko, B.: Taksonomia celow wychowania. In: Kwartalnik pedagogiczny, 24, 1979, n.2 pp. 67-78. (in Polish)

[10] Urbancik M. Management, Automation and Checking of Semestral Works. Diploma work. DCI FEEI Technical University of Kosice, 2005 (in Slovak)

[11] Genci Jan: Knowledge Assessment – Practical Example, ICETA 2008, Stara Lesna, 11.-13.09.2008, Kosice, elfa, s.r.o., 2008, pp. 153-156, ISBN 978-80-8086-089-9

[12] Telepovska Henrieta: SQL Statement Knowledge Assesment, 6th International Conference on Emerging e-Learning Technologies and Applications, Conference Proceedings, Information and Communications Technologies in Learning, Stara Lesna, September 11 – 13, 2008, Kosice, elfa, s.r.o, 2008, pp. 181-184, ISBN 978-80-8086-089-9

Diplek: an Open Educational Platform that Supports the Composition of Learning, Management and Communication Services

C. T. Rodosthenous[a], A. D. Kameas[b], P. E. Pintelas[a]
[a]Department of Mathematics, University of Patras, Greece, [b]Hellenic Open University, Patras, Greece
chrodos@upatras.gr, kameas@eap.gr, pintelas@upatras.gr

Abstract – **This paper presents Diplek, a Learning Management System (LMS) that addresses the needs of novice users, educators and students. The development of Diplek is based on a service oriented architecture. Among the many services offered by the platform, one can distinguish the innovative Video recording service and the Course creation service that enable instructors to import any type of Learning Object and associate it with metadata. Diplek offers a friendly user interface with dialogs and wizards that help users deal with difficult and time consuming tasks. The platforms architecture (not web-based) enables it to be deployed in both internet and internet-less communities.**

I. INTRODUCTION

A Learning Management System (LMS) is a software program or integrated platform that contains a series of services and tools to support a number of activities and course management procedures [1]. Nowadays numerous LMS are available in the market; a few more, like Moodle, Sakai, Atutor, Claroline etc., are under development by the open source community. Despite the many tools and services offered by the LMS, a number of limitations and disadvantages were reported from users and researchers:

1)The platform complexity and difficulty of use require the continuous Information technology (IT) support and hence cause a significant increase of total spending in new technologies [2].

2)Lack of tools for proper monitoring the students' activities throughout the session and course duration [3].

3)The dependence on internet and web technologies most of the platforms have, obstruct the deployment of distance learning services to internet-less communities and institutions

A variety of LMS/CMS have already been adopted by educational institutes worldwide; some prefer free and open source solutions and some others rely on proprietary software solutions that come with guaranteed support and helpdesk. The need for standardization led a lot of community developers to pursue a way to standardize most of the LMS services. The results of these efforts are reflected in : a) standards of *e*-Learning architecture (e.g. the IEEE LTSA, Learning Technologies Standard Architecture), b) description of learning objects meta-data based on shareable XML-based data structures (e.g. through the IEEE LOM specification) and c) the assessment and evaluation of user performance (e.g. through the IMS QTI, Question and Testing Interoperability Schemas). The above specifications enable the common description of learning units, questions and tests, learner profiles, etc, so that they can be easily interchanged between different applications [4].

A further study of the above specifications led to four main categories of services a LMS should offer to its users. These categories include services for:

- Communication between learners and instructors
- Adaptation - Personalization - Extensibility
- People Grouping and General course Coordination
- Monitoring learners' achievements and progress during a course

The purpose of this paper is to describe the design and services of Diplek, an educational platform that uses to support the needs of educators with reduced IT competence throughout the main phases of course management lifecycle. Most LMS simplify only the services that relate to the content management process; nevertheless, educators have additional needs, such as, to monitor a student's progress with means that can be easily handled by a typical non IT specialist teacher and to communicate with students in real time. Diplek offers a special tool for monitoring purposes, which records a student's learning session in video format; this recording can be interpreted at a later time by the educator to extract conclusions about the student's progress and the course's contribution in achieving its purpose. Moreover, the system can be operated without an internet connection or a web browser. This flexibility comes in handy in situations where the equipment is old and the connection between workstations is limited to a LAN. Services that require internet connectivity like "My mail" and "Live chat" are automatically switched to LAN functionally so that communication between users can be conducted normally at local level.

This paper is structured as follows. In the next section, the architecture of Diplek is described. Then, the way the system encompasses Learning Objects and the other services it offers to its users are discussed. Finally, the evaluation services are presented in a separate section.

II. DIPLEK PLATFORM ARCHITECTURE

The Diplek platform adopts a 3-tier architecture "Fig. 1,". In order to simplify the overall process, the platform operation is

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_59, © Springer Science+Business Media B.V. 2010

composed of group of services offered by several autonomous components that cooperate with each other and give the user the look and feel of a modern classroom. Users type their credentials and authentication service processes them. Each user role has a different access level; users are classified into students, educators and administrators.

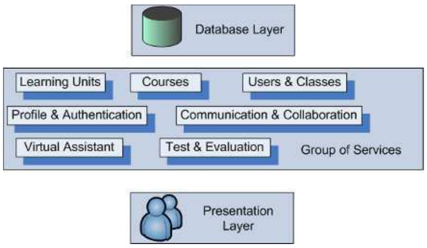


Figure 1- Layer Schema

A. Platform installation and deployment

Diplek is intended to be installed in both organizational-educational environments and home environments. In the first case, a server is needed to host the platform's database where the educational material and user data are stored. This database can be made available through the internet so that the client workstations can connect to it from any point in the world. This kind of installation is suitable for schools and universities, where students are divided into classes and each class is assigned to a group of educators. Each user (student & educator) can connect to the platform through a local area network or through the Internet. Each client workstation needs to have a Diplek client installed in order for the platform to function properly.

In the second case, where Diplek is installed in a home environment, the users' workstation acts both as a client and a server. That means that the platform is installed locally and can serve users connected to that workstation only. This kind of installation is suitable for users who cannot gain access to a permanent network connectivity.

Diplek comes with an automated configuration utility that helps the system administrator to perform easy platform installation.

B. Diplek Layers

Diplek is using a service oriented architecture [5]. This design enables the transparent addition of extra functions after the platform is deployed thus offering users new tools and services.

The services layer is the connecting link between the database and the presentation layer. Before content is displayed to the user, it's filtered and rearranged by the different group of services that intervene.

Platform services are made available to users according to the user level they have. The authentication service is responsible for distinguishing which user runs what service and sets the level of functionality of the service.

III. Learning Objects & Metadata

A Learning Object is an independent and self-standing unit of learning content that is predisposed to reuse in multiple instructional context [6]. One of Diplek's key functions is the storage and cataloging of educational material. Instructors can import Learning Objects into Diplek's main repository and accompany them with metadata.

The capability to associate metadata with Learning Objects makes Diplek a powerful tool for easy search and indexing. Metadata [7] are conventionally defined as data about data, information about information, and are used to describe document contents and structure, and to provide information about accessibility, organization of data, relations among data items, and the properties of the corresponding data domains.

Diplek uses IEEE LOM standard for its learning object repository. The reason for choosing LOM among many other well defined and documented standards is that it fully describes both the educational and technical aspect of a Learning Object. Although LOM contains a large number of attributes, Diplek requires only a small set of basic attributes to be inserted in order to allow a Learning Object to be catalogued. In this way, educators are not obliged to fill the complete form and hence can save time. The use of IEEE LOM makes Diplek Learning Object repository easy transferable, interoperable and searchable.

A. Importing Educational objects into the platform

Before creating a course, educators need to insert educational material to the learning object repository. Diplek provides an easy to complete wizard for novice users, which is divided into four steps according to the categories of metadata schema; general, life cycle, educational and technical attributes. After completing the wizard the final step is the inserting of the Learning Object. Educators can insert any type of educational material, like multimedia presentations, sounds, WebPages, applications, archives, games e.t.c . Diplek accepts all file formats. One of the main advantages of the platform is the ability to import any kind of material or even software, as long as the appropriate viewer is installed on the client's machine. In this way, all educational material can be used without any modifications or limitations.

Diplek is not an educational material creation software; it is a platform that utilizes all the available educational material in order to present it to the trainees or students. So importing the educational material is just the first step of creating a course for learners to use it. The second step is finding the appropriate material that will be presented to learners. To do so, a search tool is available that queries the platform's learning object repository using criteria based on learning object metadata (type, difficulty, level of education, language, keywords, topic e.t.c.). All Learning Objects that match the required criteria can be selected from the list. When a Learning Object is selected, it is copied in a new list. This list contains all the candidate educational material to be used for the course presentations.

B. Creating powerful presentation from Learning Objects

Course presentation is the final step of preparing a course. These presentations must not be confused with the usual slide show presentations. A course presentation is an organized structure of learning units categorized in such a way so as to serve the group's educational needs. All educational units used to form the presentation can be accessed by double clicking the mouse on their icon. These presentations are assigned to groups or specific students. The ability to distinguish what each user is viewing-using is a vast advantage since the educator has the ability to alter a presentation/course according to each user special needs.

Diplek supports the creation of course presentations with an easy to complete wizard.

First, educators fill up the general course details like title, description e.t.c. These details are shown to the student as an introduction text when the presentation is selected.

Next, follows the creation of learning units. Each learning unit is a set of Learning Objects. So, a learning unit can be a book chapter, a lab experiment, e.t.c . It's up to the educator to choose how to present a course. In this form, educators simply define the learning units by entering the title of their choice

The final stage is the distribution of Learning Objects to the corresponding learning units. A Learning Object can be used in one or more learning units or presentations. Each educational object assigned in a presentation is represented by an icon "Fig. 2,". Depending on the type of object, each icon has a different image (a generic icon is used to represent object types not recognized by the filesystem).

Figure 2- Main Presentation Screen

A presentation can be deleted or reused. Reusing a presentation can take many forms. One can reuse a presentation to teach the same course or alter the presentation to fit a different educational domain. The presentation created by an educator can be used by other educators as is or altered. Learning Objects that are copyrighted are protected since copyright information is stored in the LOM metadata set.

IV. OTHER SERVICES OFFERED TO USERS

User services relate to platform services available to all course users but which are not related specifically to course learning material (e.g., messaging between learners/ instructors, calendar, live chat, and document exchange services etc) [8].

Diplek includes a number of tools to enable learners with organizing time, organizing personal workspace, checking course grades e.t.c. These services run as components and can be configured by the instructor. All tools use a common UI and use simple command buttons to perform an action. The design of these tools was made taking into account the difficulties a novice user faces when using them. The following tools are meant to be used both by learners and instructors.

- Profile Manager
- My Calendar
- Sketch book
- My Grades (GradeBook)
- My documents (e-portfolio)
- My mail
- Live chat

Even though, the most frequently used communication tool is e-mail and it is mainly used for personal correspondence among the students [9], there are other communication tools like instant messaging, live chat and discussion forums. Diplek offers three kinds of communication tools: integrated messaging tool, live chat and email.

A. My Calendar

Calendar is a powerful feature that allows students and faculty to manage both academic and personal events. A learner must be given the ability to schedule lectures, tests, assignments and other learning processes during a week/month/year.

Instructors need to mark important dates and schedule learning activities. This tool comes in handy in case of the student users who can schedule their lectures, tests and assignment delivery. The calendar service is directly connected to the virtual assistants' services to inform user with upcoming events.

B. My Documents

Diplek supports personalization of learning and encourages users to have their own personal workspace where they can store files. These files can be anything from a single text file to an interactive multimedia game. All documents (files) can be extracted with a single click of a mouse and can be transferred to another user of the platform or even a student or instructor of another system.

C. Sketch book

Another tool that follows the learning personalization logic is the "sketch book". The use of this tool provides user with a place to enter information found from different sources (internet, educational objects, chat e.t.c) and stores them in a container that is always available during a session. Learners can search information from the web and insert it to the sketch book. This information can be saved as formatted text combined with images, and other multimedia features like sound, video, animation e.t.c. The layout of the window resembles that of an exercise book and all actions can be performed either from the top menu or by right clicking the mouse on a word or other text.

Diplek provides several other useful tools that can be used to assist learners like a calculator, a drawing tool and a web browser for internet access. Another category of tools provided by the platform are the collaborative-communication tools.

D. Live chat

Live Chat (or instantaneous mail) allows a real time discussion between all members of the platform. This is an exciting way for students to communicate directly with each other in real time, and a unique way for instructors to hold office hours. Live chat discussions can be archived for later review. Most of the people are familiar with the terms chat and internet channels. For those that are not, live chat is a tool used to simulate a room with people talking ”Fig. 3,”. Diplek does not require complex setup of chat space and the tool is ready for use when a user selects the chat option. All conversations in the room are public so that everyone can watch and participate.

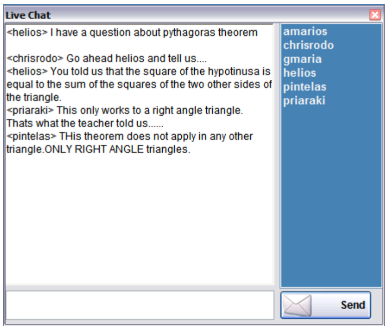


Figure 3- Live chat Window

A teacher can use this tool to give guidance to students from distance while they are watching a presentation or working on a project. This way, all students get the same information and can ask questions that everyone can see. The use of this tool is needed in order to build a collaborative community that each user learns from the other.

This tool is not recommended to be used for personal messages since every one that is in the room can see the message. For that purpose another tool can be used called ‘My Mail’. This tool is a simple messaging system for all the users that have access to the platform.

E. My Mail

A less exciting but still as useful as the live chat tool, is the ‘My mail’ service. This tool is needed for users to communicate with each other in private. The way this tool works is similar to that of the email service, but instead of email addresses, users only have to know the username of the user that they want to communicate with. Also, this tool does not require the existence of an internet connection since it uses the available LAN to deliver messages to users.

All communication logs are available to the user through a centralized system. This system gives the ability to the user to delete, create and check for new messages. A teacher has the ability to send a message to the whole classroom or to a specific user whereas student has only the ability to send a message to a specific user (student or teacher).

V. STUDENT EVALUATION AND ASSESSMENT

Byers [10] describes interactive assessment in learning environments as promoting dynamic feedback and course corrections on the fly. This part of the platform measures student performance against specified goals, using a variety of services ranging from multiple choice questions to complex assignment handling

There are many ways for evaluating a student’s progress. The most common are tests, projects and assignments. In order to cover most of these evaluation methods, Diplek includes tools that automate the above procedures. An easy and intuitive wizard-based method is introduced in order for instructors to use them.

A. Multiple choice tests

Most LMS provide templates for multiple-choice questions, true/false questions, matching questions, or short answer questions [11]. Multiple choice tests ”Fig. 4,” are an easy way of assessment and evaluation. Nevertheless, it is also a way for learners to check their learning progress. To manage the creation of multiple choice tests, Diplek incorporates services for question and test creation, test delivery and test result analysis. Instructors can create a number of questions with the desired set of answers. Each question can be given a difficulty level and can be included in one or more tests and if the instructor permits it, it can be shared by other instructors on the platform. When the test is ready, it can be assigned to a student or a group. Multiple choice tests can be used both for assessment and self evaluation.

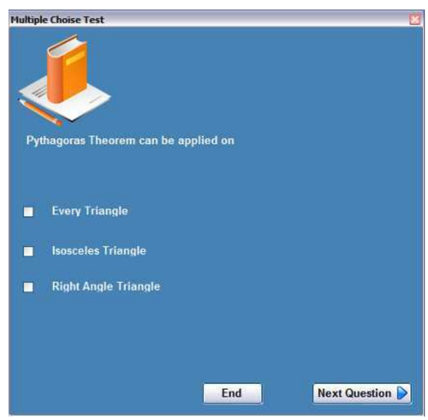


Figure 4- Multiple Choice Tests

When a test is finished, a log file is created. Instructors can view usage information by the use of the test results analyze tool so that they can determine if a test is suitable for a group or a student and make the necessary changes so that it fits the learning goals.

B. Project - Assignment Management System

Another method used for assessing a student’s performance is the project assignment service. This is a very common method used in many educational institutions. Students are involved in a search and learn procedure guided by their instructor. At the end of the day they present a document with their findings. Diplek supports this assessment method with the use of the project management service. This service can be used for the creation of an assignment, delivery and evaluation.

Teachers can create or import projects through a user friendly wizard, where they are asked to insert the guidelines, instructions to students, relevant files and deadline. After that, they can assign the selected projects to groups or individuals. Students are notified for the new assignment by either the teaching or the virtual assistant service. When the project finishes, students deliver it to the teacher by using the platform delivery tool. Teachers evaluate the project and mark it as checked. All checked projects are available to both the student and the teacher with added comments and corrections.

C. Grading system

'My Grades' tool provides the educator a list of all students attending a course where he/she can assign a grade to each one of them. Student's have access to their grades and can see them as a list. Sometimes it is more convenient to just have a printed list of all students and their grades to be published in the announcement board or delivered as an email. 'My grades' tool gives teachers the option to do that with an export tool that saves a list of grades as a spreadsheet.

Teachers can get statistics for each group or course regarding their grades assigned to students enrolled in the specific class or course. Statistical data are separated to grades per course and grades per group.

VI. USER GUIDANCE AND SUPPORT

Diplek offers besides the standard online and printed manual documentation, two additional ways of helping its users.

Teaching & Virtual Assistants

The large number of settings and services offered by the platform may prove counterproductive for the user who just wants to do some simple tasks. That's why Diplek incorporates such a tool called the 'Teaching Assistant', which is available to all students and educators.

This assistant provides information to the user about new messages, new project assignments and things to do from the calendar. User can enable and disable this feature with the click of a button so that it doesn't take space when not needed.

In addition to the Teaching Assistant, a second assistant that has a friendlier look is available to the student user. The Virtual Assistant has the form of a wizard and shows up after a successful login. It provides information about date and time of day and informs the user about new messages, projects or tests assigned to him. This information comes with instructions of how to get access to the appropriate item so that user won't have to ask how to do that.

A. Session recording service

LMS gather large logs of data of student activities during courses and usually have built-in monitoring features that enable the instructors to view some statistical data, such as a student's frequency of login, time taken on a course or a test, the number of messages the student has read or sent, marks achieved in tests, etc.

Instructors may use this information to monitor the student's progress and to identify potential problems. However, tracking data is usually provided in a tabular format, is often incomprehensible, with a poor logical organization, and is difficult to follow. As a result, Web log data is used by only few skilled and technically advanced distance learning instructors [3].

The session recording service is an innovative approach to logging students progress and learning achievements through a session. The use of this service provides the teacher with video feedback of a student's session. This feedback consists of all student actions, conversations, mouse movements, services used e.t.c. All these are recorded and saved in a video file which can be played in any video player.

Educators can view these recordings on a daily, weekly or monthly basis. By analyzing each students file, a general conclusion can be exported about the student's progress. This procedure brings a great help to the early diagnose of some of the educational difficulties that occur to students.

Session recording service can be activated for selected users or groups and is by default deactivated since it consumes a lot of system resources and is not always needed.

VII. CONCLUSIONS

Diplek is a service oriented LMS that was designed to help novice educators to get involved in e-learning technologies. There is a plethora of LMS out in the marker that can be used with a lot of great features. Diplek is not trying to compete with other LMS. More likely Diplek is intended to be used in smaller educational domains where computer experts are hard to find and the need for an easy to use LMS is prominent. The use of Diplek is recommended in situations where a continuous monitoring is needed i.e Students with educational difficulties and learning disorders.

Diplek will be distributed under an open-source license by summer 2009.

REFERENCES

[1] A. Severson, "Faculty support required for the implementation of a new learning management system.", *Thesis submitted to the Simon Fraser University for the degree of Master in Distributed Learning*, 2004

[2] L.J. Chang, C.Y. Chou, Z.H. Chen and T.W. Chan, "Strategy for developing physical and network hybrid community-based learning environment in schools: The Taiwanese experience", *International Journal of Educational Development 24* (4), pp. 361–381, 2004.

[3] R. Mazza and V. Dimitrova, "CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses", *International Journal of Human-Computer Studies*, Volume 65, Issue 2, Pages 125-139, February 2007.

[4] D. Sampson, C. Karagiannidis and F. Cardinali: "An Architecture for Web-Based e-Learning Promoting Re-usable Adaptive Educational e-Content", *Educational Technology & Society Journal of International Forum of Educational Technology & Society and IEEE Computer Society Learning Technology Task Force*, ISSN 1436-4522, Special Issue on Innovations in Learning Technologies, 5(4), August 2002.

[5] W. Shen, Q. Hao, S. Wang, Y. Li and H. Ghenniwa, "An agent-based service-oriented integration architecture for collaborative intelligent

manufacturing", *Advanced Engineering Informatics*, Volume 22, Issue 3, July 2008, Pages 328-339
[6] P. R. Polsani, "Use and abuse of reusable learning objects", *Journal of Digital Information*, 2003.
[7] M. Day, "Metadata in a nutshell", *Information Europe*, 6, 2., 2001
[8] M. S. O'Droma, I. Ganchev, F. McDonnell, "Architectural and functional design and evaluation of e-learning VUIS based on the proposed IEEE LTSA reference model", *Internet and Higher Education 6*, pp. 263–276, 2003.
[9] Y. Vovides, S. Sanchez-Alonso, V. Mitropoulou, G. Nickmans, "The use of e-learning course management systems to support learning strategies and to improve self-regulated learning", *Educational Research Review 2*, pp. 64–74, 2007.
[10] C. Byers, "Interactive assessment: An approach to enhance teaching and learning", *Journal of Interactive Learning Research*, 12(4), pp. 359–374, 2001.
[11] T. Govindasamy, "Successful implementation of e-Learning Pedagogical considerations", *The Internet and Higher Education,* Volume 4, Issues 3-4, Pages 287-299, 2002.

Double Degrees: Concerns Regarding Overall Standards and Graduate Attributes such as Probabilistic Reasoning

Bruce Moulton
Faculty of Engineering and Information Technology
University of Technology, Sydney, Australia
brucem@eng.uts.edu.au

***Abstract*—Double degrees (also known as combined degrees) typically allow engineering students to complete, in just 5 years, two degrees that would ordinarily take 7 years to complete. This paper provides and discusses the results of a pilot study relating to engineering double degrees. The study participants offered opinions relating to non-engineering fields of interest, and why they enrolled in their current double degree course. The study found that double degree students appear to be interested in breadth, but *not depth*. This finding seems to contradict a prevailing view that double degrees offer students the potential to gain depth in a particular niche or "overlap area" between the two degrees. The paper also discusses double degree curricula, in particular, the subjects that are "omitted" from a typical double degree. It is noted that some elements that would ordinarily be strongly associated with depth and critical thinking seem to be missing from double degrees. The issue is exemplified in this paper by focusing on one aspect of critical thinking: probabilistic reasoning.**

***Index Terms*—engineering education, critical thinking, judgment and decision making probabilistic reasoning, double degree, combined degree**

I. Introduction

INDUCTIVE reasoning may be our most important and most prevalent problem-solving activity [1]. The term "probabilistic reasoning" refers to sound inductive reasoning that relies on probabilistic principles. Examples include reasoning that considers the number of instances (sample size), whether a sample is unbiased, and whether instances concern events of low or high variability. "Generic" critical thinking skills such as probabilistic reasoning are recognized as some of the most important skills that higher education courses can enable.

While "single" degrees are generally built on a carefully designed and balanced program that seeks to balance breadth, depth, generic and specialist skills, this is not the case with double degrees. Double degree programs are typically the result of a quick and dirty process where a substantial amount of material is removed from two completely different degrees, and what remains is then bolted together. This has led some commentators to offer views to the effect that double degrees are "not interdisciplinary and barely integrated" [2: 149].

The term "double degree" is understood in different ways, and varies from one country to another (and in some cases, differs between universities of the same country). In general, it seems a double degree is a particular type of *cross disciplinary* educational program where students complete two degrees, undergraduate and/or postgraduate, either consecutively or simultaneously, through two faculties or disciplines of one or two institutions. In this paper, the term "double degree" mainly refers to the first two cross-disciplinary educational models described in Table 1.

TABLE 1
VARIOUS DOUBLE DEGREE MODELS

Model	Label
Two undergraduate degrees studied simultaneously through two faculties of one institution	Combined Degree (Aus), Conjoint Degree (NZ), Double Degree (Aus), Dual Degree (Aus, USA), Joint Degree (UK, USA)
Two degrees studied simultaneously at two different institutions, sometimes internationally	3:2 Program (USA), Coordinated Degree (USA), Double Degree (Aus), Dual Degree (UK), Joint Degree (Aus)
Two degrees studied consecutively at one institution	Articulated Degree (Aus) Integrated Degree (Aus), e.g. Uni of Melbourne 3+2 model
Two degrees studied consecutively at two institutions in two different countries	Consecutive Degree (Coimbra Group), Dual Degree (UK), Joint Degree (Aus)
One degree offered by two faculties of one institution	Joint degree (Aus)

Engineering-science double degrees are generally thought to build capacity for highly demanding technical and research oriented areas of engineering [3]. Double degrees that combine engineering with arts, business, commerce and management have gained in popularity in the last 2 decades, and have proven particularly attractive to women – these degrees are presumed to provide students and graduates with broader study and career options.

Bruce Moulton is a Senior Lecturer and Head of Combined Degrees Program with the Faculty of Engineering and Information Technology, University of Technology Sydney. Phone: 61 2 9514 2655; e-mail: brucem@eng.uts.edu.au

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_60,

For disciplines such as science and engineering, there is a perception that double degree programs retain students who might otherwise have solely undertaken studies in business or management.

The popularity of double degrees in engineering suggests that students perceive double degrees to be beneficial. However, such beliefs may be unsubstantiated – outcomes of completing an engineering double degree appear to be largely unknown. Notwithstanding, even if it were found that graduates of double degrees appear to perform better than those of single degrees, such findings may be confounded by factors such as the higher entry requirements of double degrees.

A general principle of constructivist learning theory is that people learn by building on their current knowledge. Amongst other things, the theory predicts that, all other things being equal, learning activities designed with an understanding of the learners' current attributes would be more effective than learning activities that disregard learners' current attributes. While considerable effort has already being expended to improve our understanding of the current attributes and interests of students of "standard" degrees, the same cannot be said of double degrees. Students of double degrees may have many features in common with their standard-degree counterparts. However, it is quite possible that double degree students may differ in certain ways– for one thing, they are willing to wait longer until they graduate! It seems plausible that some double degree students might be more interested in attaining greater *breadth*. It is also possible that some might be interested in greater *depth,* perhaps in particular niche areas such as "engineering management".

Other questions regarding double degrees relate to the coverage of the programs, whether the programs are meeting expectations, and whether the material that is "omitted" from each degree detrimentally affects the development of certain graduate attributes.

II. Method

In March 2008, 16 engineering double degree students were surveyed about their perceptions of double degrees. The students volunteered to stay back after a lecture to complete a one page survey regarding double degrees. The survey consisted of 6 short-answer questions, and was designed to learn about students' interest in double degrees, interests in transferring from once course to another, and interests in other fields and activities. We are mainly concerned here with their answers to questions relating to the factors were relevant for students' decisions to enroll in their current courses.

III. Results

The comments of the surveyed students were analyzed and characterized according to whether they seemed to relate to breadth or depth. During this process it became apparent that none of the comments seemed to relate to the adopted concept of depth.

The results suggest that students who undertake double degrees are in some cases attempting to balance personal interests (such as science) with career interests (such as business); other perceptions include "mix it up" and "because it's cool. Representative comments are provided in Table 2

TABLE 2
Representative comments regarding each theme

Theme	Representative comments
"Breadth" / career	I wanted a career in both science and engineering these two compliment each other
	Keep options open. Interest in finance and engineering. Did not want to be purely designer, more interested in management, construction etc
	BEBsci, BEBBiotech, I feel it could open up many other avenues that could prove to break up the engineering degree
	I am interested in doing technology consulting and a business degree would provide me with an understanding of businesses and their points of view. I am also interested in economics
	Qualifications in two fields - fields are related - biomedical engineering
Interests v career	Business re job, science re interests..
	Career objectives, academic interest (BE-MedSci student)
	Science as I was always interested in some science subject. Perhaps business to give a better job opportunity
Other	Opportunity to lean a new language and experience a new culture. University offers support and help to go overseas to partner uni (BE-BA student)
	Fun, mix it up, overseas (BE-BA student)
	Because it's cool (BE-BA student)
Depth	None

IV. Discussion

A. There appears to be a perception of breadth, but not depth, nor integration between the two courses

It was somewhat surprising that no comments relating to depth were found. While some students alluded to niche areas, these seemed more centered on "covering two fields" rather than achieving depth. To some extent, this seems to contradict the prevailing view that double degrees enable students to achieve depth in a niche area at the overlap of the two degrees. Double degrees are commonly thought of as a means to develop highly specialized niche areas such as science-law and engineering-management. Such integrated niches are thought to be valued by employers. It was thought that students and employers might assume double degree curricula in such niche areas would include depth at the interface of the two degrees.

For example, it might be expected that a course such as BEng-BMedSci (Bachelor of Engineering Bachelor of Medical Science double degree) would enable graduate attributes in the area of biomedical engineering. However, in reality, the course may simply include 'standard' engineering alongside 'standard' medical science. Thus it is proposed that double degree curriculum development and renewal processes, particularly in relation to depth at the interfaces of disciplines, may be below expectations.

A 1996 review by The Institution of Engineers Australia, known as *Changing the Culture,* encouraged schools of engineering to develop curricula that would enable graduates to work across disciplines. Twelve years later, the 2008 ACED review expresses the view that with respect to work across the disciplines, there has only been a limited degree of progress. The review finds "the undergraduate programs in the two disciplines have tended to remain completely separate" ([3:81), and that although engineering practice regularly takes place in business environments, apart from some coverage relating to introductory management, there appear to be few curriculum links with management schools. The review also finds that there are not many opportunities for students of engineering-management double degrees to undertake projects that explicitly take advantage of the combined disciplinary coverage.

A study by Russell, Dolnicar & Ayoub [4] found only a small number of students identified the skills learnt in their double degree as useful at work, even though it appears that many students took double degrees to increase employment opportunities or to explore new areas of professional interest. Those findings seem broadly consistent with some of the responses in this study.

B. *Learning critical thinking / probabilistic reasoning*

Another concern regarding double degrees is the extent to which they enable generic skills and critical thinking. A central aspect of critical thinking is inductive/probabilistic reasoning. We focus here on probabilistic reasoning as an example of one of the generic skills that may be insufficiently addressed in double degrees. By way of explanation, a seminal study concerned with probabilistic reasoning included the following question:

> A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital, about 15... As you know, about 50% of all babies are boys. The exact percentage of baby boys, however, varies from day to day. Sometimes it may be higher than 50%, sometimes lower. For a period of 1 year each hospital recorded the days on which (more/less) than 60% of the babies born were boys. Which hospital do you think recorded more such days? ([5], p. 443).

Most of Kahneman and Tversky's participants answered that they thought days with 60% or more male births would be equally common at the two hospitals. Their findings indicated a lack of probabilistic reasoning, as such "highly deviant" days should be less common when samples are larger (in this case, at the hospital with more births per day).

Nisbett et al. [1] conducted a study where participants with differing levels of statistical education were asked to explain circumstances such as

> Why might someone experience a truly outstanding meal on a first visit to a restaurant but then be disappointed on a repeat visit? (p. 358).

Nisbett et al. termed explanations such as "the chef might have changed" as "non statistical". "Statistical" explanations included answers such as "although many restaurants *sometimes* produce excellent meals, very few restaurants *always* produce excellent meals" [1:358]. It was concluded that participants who had greater levels of statistical training gave more statistical explanations for scenarios such as that of the disappointed customer; that is, people with greater levels of statistical training tend to exhibit a greater degree of probabilistic reasoning. The study also investigated other methods for improving probabilistic reasoning, and tested a theory that one method for improving probabilistic reasoning might be to increase the salience of statistical factors. To this end, the study, manipulated participants' recognition of the *sample-like nature* of events, and found that people do tend to reason more probabilistically when the sample-like nature of events is more apparent.

Similarly, [6] found that instruction in statistics had a significant effect on the way their participants reasoned about everyday problems, and instructions including examples were particularly effective [6:272]. They concluded educating people about concepts such as the law of large numbers improved their reasoning;

> [the studies] clearly demonstrate how classroom training in statistics can potentially have a significant effect on how people make judgments. If introductory statistics courses were to incorporate examples of how statistical principles such as the law of large numbers can be applied to judgments in everyday life, we have no doubt that such courses would have a more far-reaching effect on the extent to which people think statistically about the world [6:282].

Studies such as these suggest that decision-making judgment can be developed by training that is directed towards the development of probabilistic reasoning. In an engineering degree, probabilistic reasoning is developed in at least 3 ways.

First, it is taught within 2 or 3 mathematics subjects which teach “mathematics as mathematics”. Second, such reasoning skills are developed in advanced “sub-majors” and chains of subjects where each subject builds upon embedded mathematical knowledge developed in the subject(s) that precedes it. Obvious examples of this are seen in areas such signal processing or electrical engineering. Third, probabilistic reasoning is one of the graduate attributes that is developed through engineering internships, and double degree students typically undertake much shorter internships than single degree students.

According to [7], judgment and expertise is developed gradually over time, and best developed in “stages”; students at early stages might rely more heavily on teacher-supplied feedback, whereas students at later stages might engage in greater levels of learner self-monitoring. Such “stage” theories suggest that critical thinking skills are developed in incremental steps. First, it is usually necessary for each subject coordinator to know something about the students’ current level of development in order to pitch the learning activities at a suitable level. Second, it is necessary to ensure that curriculum developers do not alter the program in a way that results in various educational stages being omitted – this is particularly relevant for double degrees because subjects are typically deleted from each degree fairly expediently before the two degrees are slotted together.

C. Some elements of the standard courses that are strongly associated with depth and critical thinking are deleted from double degrees

A student would ordinarily require 7 years to complete a BE (Bachelor of Engineering) and a BBus (Bachelor of Business). When the two are combined as a double degree, this can be reduced to 5 years.

Another way of viewing this is to look at the total number of hours studied. At UTS, students typically study 4 subjects at a time. Each subject nominally ‘expects’ 9-12 hours of study per week. (Thus a typical full time student has a total nominal load of about 40 hours per week.) Total nominal hours studied might look as follows:

BE: 4032 hours (4 yrs)
BBus: 3024 hours (3 yrs)
The two as separate degrees: 7056 hours (7 yrs)
BE-Bus double degree: 5040 hours (5 yrs)

This raises the following question: in the double degree, what happens to the other 2 years? The answer relies in part on recognition of prior learning.

To explain, when a student enrolls in a degree, exemptions from certain subjects may be granted if the student has already undertaken prior tertiary studies. An engineering student who already has a business degree might be exempted from 2-6 subjects, which would typically include an introductory mathematics subject, an introductory engineering-economics subject and up to 4 elective subjects.

When a student undertakes a double degree, the prior learning that occurs in the each degree permits exemptions in the other degree – each degree recognizes the prior learning of the other.

For example, at UTS, when a BE is combined with a BBus, the effect is that a double degree student studies 16 less subjects than the total of the 2 separate degrees: the BE-BBus omits 7 subjects from the standard engineering degree, and 9 subjects from the standard business degree.

The 16 omitted subjects are shown in Table 3.

TABLE 3
OMITTED SUBJECTS (16 SUBJECTS, 7 FROM ENGINEERING AND 9 FROM BUSINESS

Engineering	Business
2 engineering Core subjects: Engineering Enterprise Engineering Economics and Finance	1 core business subject: Business Statistics or Managing People and Organisations
1 engineering major subject (in the case of Civil Eng, this is Surveying, Geotech or Hydrology)	
4 engineering sub-major subjects	8 Business sub-major subjects (2 sub-majors of 4 subjects each)

The 16 omitted subjects include 3 core subjects, at least 2 of which are highly mathematical, and 12 sub-major subjects, which are typically associated with specialty areas that afford students greater depth in particular areas. The overall standards of double degrees are to a large extent dependent on the standards of the single degree parts from which they are made. Thus it is important to consider the RPL-related rational that is applied in the formation of double degrees, and to investigate the extent to which graduate attributes relating to professional, generic skills, or transferable skills are omitted from double degrees.

D. Concluding remarks

Significant concerns exist regarding double degrees. Double degrees can prevent students from achieving sufficient depth in their specialized areas of interest, and this can make it harder for students to continue through to postgraduate research. Students of some double degrees are studying less mathematics. The fact that sub-majors are deleted from double degrees means that students are doing less “chains” of subjects, and those “chains” are often where more-sophisticated critical thinking and probabilistic reasoning is developed. In addition, the scant prior research to date suggests that double degrees might enable learning in two disciplines, but, paradoxically, seem to lack cross-disciplinarity. Taken together, these factors can decrease the opportunities for double degree students to continue into postgraduate research, and can detrimentally impact on the development of generic skills such as probabilistic reasoning.

References

[1] Nisbett, R.E., Krantz, D. H., Jepson, C. and Kunda, Z.. (1983). The use of statistical heuristics in everyday inductive reasoning. Psychological Review, 90, 339-363.

[2] Welsman, S.J. (2007). Double or nothing! Clever thinking, double-degree frustration, and returns to Science, URL: www.science.uniserve.edu.au/pubs/procs/2007 (Accessed 1.3.2008)

[3] King, R. (2008). Addressing the supply and quality of engineering graduates for the new century Australia: Carrick Institute for Learning and Teaching in Higher Education

[4] Russell, A., Dolnicar, S. and Ayoub M. (2008). Double degrees: double the trouble or twice the return? Higher Education, 55:575-591.

[5] Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive Psychology, 3, 430-454.

[6] Fong, G.T. Krantz, D.H. and Nisbett, R.E. (1986). The effects of statistical training on thinking about everyday problems. Cognitive Psychology, 18, 253-292.

[7] Sadler, D.R. (1989). Formative assessment and the *design of instructional* systems. Instructional Science, 18, 119-144

Pro Bono in Engineering: Towards an Improved Understanding

Bruce Moulton
Faculty of Engineering and Information Technology
University of Technology, Sydney, Australia
brucem@eng.uts.edu.au

***Abstract*—Pro bono work is more commonly associated with the legal profession than with the engineering profession. It is work that is undertaken at significantly reduced cost to a client. While many engineering professionals do undertake pro bono work, when compared with law and medicine, engineering pro bono work is not as well-known, nor does it attract high-profile support and visible recognition. This is despite the fact that the work of engineering organizations typically has great impact, for example by achieving safer water and improved sanitation. However, even though pro bono engineering work may be comparatively less visible, it is definitely occurring, frequently in ways that successfully address local capacity building, health, sanitation, information technology and housing issues among the most vulnerable and disadvantaged. This notwithstanding, it seems that a pro bono *culture* is less developed in the engineering profession than in the legal profession. It is proposed that part of the process for facilitating a pro bono culture within the engineering profession involves engineering education, and that to this end an important step is to understand engineering students' current attitudes and values with respect to the principles of pro bono. Thus this paper provides the results of a preliminary study about students' attitudes and values about pro bono engineering work. The results of this study suggest that only 10% *clearly intend to,* and a further 20% *might,* undertake some form of occasional pro bono work at some point in their careers, together with a general understanding of engineering pro bono. A goal of the study is to explore some underlying issues, with a view to making recommendations on the feasibility and viability of possible approaches for supporting greater recognition and adoption of principles of pro bono within the engineering profession. Pro bono in engineering is relatively unexplored as an area for potentially valuable research, and this paper discusses some of the avenues for potential research. It is concluded that there is great potential for mechanisms to be developed to encourage the principles of pro bono to be more effectively embraced by engineering education programs.**

***Index Terms*— development, engineering education, ethics, pro bono**

I. Introduction

PRO BONO is an abbreviation of the Latin *pro bono publico* – for the public good. The term is commonly associated with the legal profession, and refers to work that is undertaken at no cost, or at significantly reduced cost, to a client. Principles and practice of pro bono are widely recognized and adopted by legal practitioners, but seemingly, less so by engineering practitioners. For example, regarding hours of pro bono work performed, data is readily available with respect to legal practitioners – according to the Australian Bureau of Statistics, Australian solicitors and barristers undertook 2.3 million hours of pro bono work in 2001-02 [1], but such data is not available with respect to engineering practitioners. In addition, while systems that enable no or low cost access to the legal profession are widespread, such systems are comparatively less widespread within the engineering profession.

Principles of pro bono apply to much of the work undertaken by organizations such as Medecins Sans Frontieres and Engineers Without Borders. The engineering organizations are comparatively less well known, and attract less support and recognition, even though in many cases the work of engineering organizations has great health impacts, for example, by achieving safer water and improved sanitation. A large proportion of engineering aid occurs at a comparatively early point in the causal chain, closer to the *root causes* of poor health or wellbeing (as compared to medical treatment of the *symptoms or consequences* of insufficient engineering infrastructure.

Pro bono *targets* (an expected number of hours per year) currently exist in the legal profession [2] – these include the American Bar Association's Model Rule 6.1, which states a lawyer should aspire to render at least 50 hours of pro bono legal services per year [3]. Even though pro bono is by nature voluntary, some hold the view that compulsory targets are appropriate for students of law (e.g. [4]). This raises questions relating to whether similar targets might be appropriate for students of engineering.

Prior work on engineering ethics and engineering in developing countries includes [5], [6], [7], [8], [9] and [10]. However, there is very little prior research that specifically relates to pro bono in engineering.

Pro bono legal services are frequently associated with circumstances that fall within a human rights area (e.g. [11], [12], [13] and [14]). For example, pro bono legal work sometimes relates to issues arising from homelessness, or services provided by state authorities to members of indigenous groups. In addition, legal pro bono clients tend to be drawn from the most vulnerable and disadvantaged people in the community. Pro bono engineering work has some

Bruce Moulton is a Senior Lecturer with the Faculty of Engineering and Information Technology, University of Technology Sydney. Phone: 61 2 9514 2655; e-mail: brucem@eng.uts.edu.au

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_61,

similarities with its legal counterpart – for example, pro bono engineering work is frequently intended to addresses local capacity building, health, sanitation and housing issues among the most vulnerable and disadvantaged.

It is proposed that a culture of pro bono seems to be less developed in the engineering profession than in the legal profession. Some might be tempted to joke that it seems engineers may be even less ethical than lawyers. Regardless, all are answerable to the effects of widespread greed, poverty, inequality and the failure to meet basic requirements such as potable water. There is much more that the engineering profession could do to minimize needless death, disease, malnutrition and dehydration.

If we are to bring about change, we might begin by seeking to understand what engineering students "believe". A basic tenet of constructivist learning theory is that people learn by building on their current knowledge. The theory predicts, all other things being equal, that learning activities designed with an understanding of learners' current attributes would be more effective than learning activities that disregard learners' current attributes. Thus we seek to understand more about our students' attitudes and values.

A goal of this work is to explore some underlying issues, with a view to making recommendations regarding the feasibility and viability of possible approaches for supporting greater recognition and adoption of the principles of pro bono within the engineering profession.

II. Method

In September 2008, students of UTS engineering who had completed their first internship submitted a report about their learning. These are junior students, having completed only about 1.5 years of their engineering degree. Amongst other things, the report assessment criteria included the following: "Gives informed consideration of your possible future work-activities, including potential roles for you in (a) the private sector, (b) the government sector, (c) the NGO sector and (d) pro bono". For this study, 50 reports were chosen at random and analyzed.

III. Results

Approximately 20% of the reports seemed poorly engaged with the pro bono aspect of the writing task. For example, some of these reports did not indicate *informed* consideration of pro bono. Such reports typically included comments such as the following.

> *"There is very little opportunity for software engineers to perform pro bono work. Other than setting up networks and websites for schools or charities, pro bono work in the software industry is rare"*

(An enormous amount of software development/support is done at low or no cost, for example, public help forums and much of the work towards Open Source/Linux/GNU etc.)

About 10% of the students indicated that they had not considered pro bono before. An example is:

> *"Working pro bono is another path I had not considered but upon additional thought I decided that I would probably like to dedicate some of my time to helping people that can not afford the great expense of hiring a consulting engineer..*

Another 10% of students readily acknowledged that they had no knowledge, or a fairly limited knowledge, of pro bono in the engineering sector;

> *"I do not know much about the Pro bono sector for engineers. I know there are many voluntary counselors and legal teams doing this kind of service for the public. I have no intentions at the early stages of my engineering career to be working for the pro bono sector, as first of all, I do not have the adequate knowledge and skill for the jobs that might apply for me, secondly I need a solid financial base to be doing voluntary work. But on the other hand, doing voluntary work for an internship is still a valuable opportunity for gaining experience."*

About 30% of the reports were characterized as having comments that seemed to focus to a greater extent on what students could "get" out of pro bono work, as compared to how they might be able to participate in work that is intended to meet the *needs* of potential pro bono clients;

> *"Aiming for a role in the pro bono sector for me would be very unlikely. I believe working in this sector won't reap much reward and I won't learn as much as I could if I was to work in another sector."*
>
> and
>
> *"Currently I have no intention to pursue Pro Bono work. I do not have any interest in undertaking any professional, graduate or undergraduate pro bono work and as a result I will not pursue it in the immediate future."*

About 20% of the reports indicated that they *might* participate in pro bono at *some later time,* and also demonstrated a reasonable understanding of pro bono;

> *"I can see myself working pro bono for purposes I strongly believe in such as breaking down barriers by providing the less fortunate with homes and an adequate infrastructure for their community. Such work will not only assist the community in third world countries for instance, but will also allow me to grow as a human being as I become exposed to other cultures and a different way of life. Doing this can leave me refreshed and energised for when I return to the public or private sector for example, as I will be filled with innovative ideas to help the less*

> *fortunate communities. If I am to undertake work-activities pro bono, I am likely to do so after at least five years of work in either the private and/ or public sector"*

and

> *"Pro Bono is described as work done for the public good or professional work completed voluntarily. I am not sure if I will really get involved in any Pro Bono work within the foreseeable future however it may be something I would want to complete at a later stage in my career. An example of this would be working as a professional consultant in a volunteer organization based in a third world country constructing public buildings such as schools and private residences. This could be something like the research done at UTS to earthquake proof mud brick housing."*.

Only 10% of the reports demonstrated a *clear intention* to be involved in pro bono work, even if only on an occasional basis, and also demonstrated a basic understanding of pro bono as work which is undertaken and directed towards addressing client/community needs;

> *"Pro bono work, or work for the public good, is a field where I can definitely see myself occasionally helping out with. This could include helping people who wish to get their ideas from a drawing board to a prototype. I could help guide them and show them where they can make changes based on my experience."*

and

> *"I think that pro bono should be better encouraged and promoted by Engineers Australia. It is a great way to utilize your professional skills and give something back to the world without sacrificing a good income. It is definitely something I want to be involved in after university."*

IV. Discussion

The results of this study suggest that only 30% of the sample indicated that they *would* or *might* undertake some form of occasional pro bono work at some point in their careers. To some extent the results are broadly consistent with the proposition that pro bono in engineering is insufficiently recognized. The findings suggest there is significant immediate potential for mechanisms to be developed to encourage principles of pro bono to be more effectively embraced by engineering education programs.

A. Areas for further research

Pro bono in engineering is relatively unexplored as an area for potentially valuable research. Compared with law, even basic benchmarking data is not currently being recorded. Questions that remain to be addressed include the following:

- Is it feasible and viable to more-systematically support pro bono in engineering? How might aspects of the way pro bono work is recognized, supported and performed in the legal profession be applied to engineering? Research in this area might compare and contrast examples of legal work with engineering work, and consider issues such as confidentiality, insurance, intellectual property and quality assurance.

- Are some types of engineering work more suited to pro bono than others? Can chunks of some types of engineering work be readily identified as suitable for pro bono? If so, which types of work are most suitable, and which are least suitable?

- Could there be a significant role for real-time web systems about things such as (a) eliciting/codifying client needs; two modes: clients register their needs via web; the system interacts with the web and identifies needs as events/disasters occur (b) project managing/sub-contracting client needs into smaller pieces of work (c) identifying pieces of work which can be done from remote locations (online) (d) eliciting/codifying volunteer capacity and (e) matching contracts with capacity. Research along these lines might look not only at software development, but also any field where designs are exchanged and developed via the web. Such work might perhaps draw on the experience of open source software development.

- How might the principles of pro bono be more effectively embraced by engineering educators, and within engineering education programs?

- Who undertakes pro bono engineering work, and who are the clients? Questions here would include experience as professional engineers, which firms/employers are involved, and who are the clients (e.g. friends, family, community organizations)

- What pro bono engineering work is being undertaken, e.g., advice, design, development, implementation, maintenance, improvement – and in what fields, e.g., mechanical, electrical, civil, construction, computer systems, telecommunications? How much time is given?

- When is pro bono engineering work undertaken? Questions here might relate to whether the work was planned or performed or ad hoc, or during practitioner's 'vacations' or weekends.

- Where are pro bono practitioners generally located geographically, and where are their clients located?

- Why do some engineers undertake pro bono work, and why do others not? What are the motivations, barriers, relationships between engineers and clients, employer organization practices? For example, is pro bono work taken into account in salary reviews or promotion criteria, what are practitioners' values and reasons for undertaking pro bono work? Related questions would address whether practitioners believe targets should be set regarding pro bono work to be performed.

- How is engineering pro bono work currently performed? What existing approaches are successful in supporting pro bono engineering work? For example, how is the work undertaken within a structure such as those provided by organizations such as Engineers Without Borders, and does donated equipment or materials support the work?

- What are the obstacles to the consideration of voluntary and compulsory targets in engineering? (Noting that such targets have been considered and adopted in law, e.g. [15].)

B. Towards an improved understanding

Pro bono engineering work embraces extremely important problems relating to poverty, inequality and human development.

Data relating to some aspects of pro bono engineering work is virtually non-existent at present. Such data is widely available within other professions such as law, and is regularly published in a way that provides professional recognition of individual or collaborative undertakings. In law, such data has been used to enable further adoption of the principles of pro bono within the legal profession. It could be reasonably expected that the acquisition and publication of data concerning pro bono engineering work would make a significant contribution to enabling incentives and increased adoption of pro bono principles within engineering.

There may also be a role for aspirational targets in the engineering profession, in a similar way to the widely understood and adopted role of such targets within the legal profession. By exploring barriers to engineering pro bono work, and mechanisms by which the work might be better understood and facilitated, it could be reasonably expected that the research would provide information which would enable greater understanding and more-informed policy decision-making in various settings, including local communities, educational program development, and public policy recommendations.

Further research would likely contribute to an improvement in the ways that pro bono engineering work is recognized, enabled and supported. If this turned out to be the case, consequences of the research would likely have significant positive implications for not only the engineering sector, but also for the beneficiaries of additional pro bono work that is undertaken.

In addition, it is widely known within the legal profession that pro bono work, if well performed, commonly leads to non pro bono work at commercial rates. If this is also the case within engineering, improvements to the recognition and support of pro bono engineering work would likely lead to increased opportunities and competitiveness for non pro bono work.

Further engineering pro bono research would likely offer a significant contribution to discourse within the profession and the public domain, with eventual effects on engineering-related education, norms, values and perceptions. In this way, it is envisaged that further research could contribute to a more cohesive community where engineers and their clients have greater levels of participation in processes that seek to address non-engineering needs as well as engineering needs.

It is thought that a significant amount of pro bono work undertaken by engineers is undertaken with respect to infrastructure relating to issues such as water-quality, sanitation, transport and communications for remotely located communities. Given that elements of such infrastructure are known to be directly connected with disease and other health risks, it follows that efforts directed towards building local capacity with respect to preventative healthcare would benefit from further research, especially where pro bono engineering work is directed towards developing local community capacity in areas such as water quality and sanitation.

Another potential avenue for engineering pro bono research concerns the means by which to enable a collective shift in the profession beyond treating symptoms towards treating the underlying problems, for example, moving from "building homes for the homeless" towards "preventing homelessness".

If it could be achieved, a shift in consciousness would likely produce outcomes in several areas. First, it could result in improved understanding and strengthening of ties between the engineering community and the wider community. Second, it would likely have implications for professional-association guidelines and public policy decisions that could lead to improved incentives and support for pro bono engineering work. Over the longer term, it would likely lead to mechanisms that would enable improved communication infrastructure, improved local engineering capacity and improved access to technical assistance. In addition, there is the potential for improved participation in public discourse and public policy.

Acknowledgment

The author acknowledges contributions made by Alex McGuirk, Josh Barr and Robin Braun.

References

[1] Australian Bureau of Statistics (2003), 8667.0, Legal Practices, Australia 2001-02 available at http://www.abs.gov.au/

[2] Evans, A. (2003) "Recognising the conditional nature of pro bono motivation: avoiding 'aspirational' compulsion and developing an appropriate pro bono ethic in new

lawyers" Second National Pro Bono Conference: Transforming Access to Justice, Sydney, 20-21 October 2003

[3] American Bar Association (2002) Model Rules of Professional Conduct Public Service Rule 6.1 Voluntary Pro Bono Publico Service, accessed 5/1/08, http://www.abanet.org/cpr/mrpc/rule_6_1.html

[4] Cole, M. (2003) "Are compulsory targets in the public interest?" Second National Pro Bono Conference: Transforming Access to Justice, Sydney, 20-21 October 2003

[5] Florman, S. (1987) The Civilized Engineer, St Martin's Press: New York

[6] Florman, S. (1996) The Introspective Engineer, St Martin's Press: New York

[7] Harpham, T. and Anelay, L. (1999) "After roads and dams: what role for engineers in the poverty reduction strategies of bilateral agencies?" in Journal of International Development, vol. 11, issue 6, 811-823

[8] Beder, S., 1995, Engineers, ethics and sustainable development, Tenth International Congress of Logic, Methodology and Philosophy of Science, Florence.

[9] Johnston, S. F., Gostelow, J. P. and Jones, E., 1999, Engineering and Society: An Australian Perspective, 2nd Ed (South Melbourne: Addison Wesley Longman)

[10] Martin, M. W. and Schinzinger, R., 1996, Ethics in Engineering, 3rd edn (New York: McGraw-Hill).

[11] Pogge, T. (2005) Human Rights and Global Health: A Research Program. *Metaphilosophy*, 36, 182-209

[12] Pogge, T (2007) 'Severe Poverty as a Human Rights Violation,' in *Freedom from Poverty as Human Right: Who Owes What to the Very Poor*. Oxford: Oxford University Press 2007.

[13] National Pro Bono Resource Centre. *Transforming Access to Justice: Second National Pro Bono Conference Issues and Ideas*, National Pro Bono Resource Centre, UNSW, Sydney 2005

[14] National Pro Bono Resource Centre. *Mapping pro bono in Australia.* Kensington, NSW, National Pro Bono Resource Centre, 2007

[15] Hillard, D. "Establishing a voluntary minimum pro bono target for Australian lawyers" Second National Pro Bono Conference: Transforming Access to Justice, Sydney, 20-21 October 2003

Experiences from Software Maintenance Seminars: Organizing Three Seminars with 127 Groups

Jussi Koskinen
Department of Computer Science and Information Systems, University of Jyväskylä,
P.O. Box 35, 40014 Jyväskylä, Finland, koskinen@jyu.fi

***Abstract*-Software maintenance and evolution (SME) is an important but problematic topic-area for university-level computer science education. Seminars can be used to provide versatile and up-to-date knowledge for students regarding scientifically relevant issues. We have organized three systematic university-level seminars on SME with a total of 127 seminar groups. Each group has been assigned a task of analyzing one scientific SME-article. The main results include the general confirmed feasibility of the selected seminar-based approach. The paper describes the background of the seminars, their contents, and experiences concerning organization and feasibility of the seminars. The results support organizing, studying, and improving feasible seminars in software engineering.**

I. INTRODUCTION

Software maintenance and evolution (SME) is a large, rapidly advancing, and economically important field. The relative amount of the costs of SME-activities has traditionally been 50-75% of the total software life-cycle costs. Newer studies such as [1] show similar or even higher proportional costs. Especially maintenance of poorly documented legacy software systems is often very elaborate but necessary.

Due to these underlying reasons, there is a constant need in computer science education to follow the scientific progress, and transfer the achieved results to future professionals of this area. This is important also in order to preserve and promote student interest on the area. Software maintenance has traditionally not been among the most fashionable subject areas among students, companies or academic staff.

Therefore, there is a need to experiment with new forms of teaching. Organization of seminars is one promising option for increasing both interactivity, and scope of the subject areas which can be covered in academic teaching. There are some scientific studies on SME-courses, but surprisingly no earlier systematic studies on SME-seminars. Therefore, there is a clear need to study the feasibility, potential and limitations of seminars as an approach to introduce SME-related knowledge.

This paper describes experiences from three software maintenance seminars with 127 seminar groups and 296 involved students. The paper is organized as follows. Section II is an introduction to the general state, problems and options of teaching SME. Section III briefly describes the organized seminars. Section IV presents experiences based on the organized seminars. Finally, Section V concludes the paper.

II. TEACHING SME

We will first describe the general context of teaching SME (*A*), the main problems related to it (*B*), and the consequent general need for organizing and studying SME-seminars (*C*).

A. The general context of teaching SME

A.1. Teaching software engineering. The IEEE's project SWEBOK has suggested an informed and reasonable characterization of the software engineering body of knowledge [2]. It covers the field of software engineering extensively. Among other things it is stated to form a foundation for curricula development. That knowledge has been used in the joint ventures of IEEE and ACM producing software engineering curriculum guidelines [3,4]. Software maintenance is one of the 11 knowledge areas of SWEBOK. Important SME-issues include maintenance processes and techniques, management issues, and software evolution.

A.2. Teaching SME. Software engineering curricula development has also been discussed in general level e.g. in [5,6] including some remarks on SME. There are also various descriptions of courses on SME in university-level published in scientific series. However, none of the studies has dealt with seminars concerning scientific articles. Therefore, there is only a limited amount of directly related studies. Most of the earlier studies have focused either on general course descriptions or on bringing forth practical sides of SME.

A.3. Teaching SME at the University of Jyväskylä (UJ). We have considered the state of teaching SME and its problems at UJ, which provides conventional university-level computer science education leading to candidate's degree, master's degree and doctoral degree. Teaching of SME in the Faculty of Information Technology (FIT) of UJ has mainly been organized by the author of this paper. Currently, there are 15 credit units of advanced-level non-mandatory SME-courses provided for students. One credit unit corresponds to 40 student work hours. FIT has systematically aimed at following in general level the recommendations of IEEE and ACM [3,4] since 2002, concerning the composition of the courses it provides, e.g. regarding the coverage and emphasis of essential issues. Based on these activities we have identified the following three main practical problems in teaching SME.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_62,

B. Problems of teaching SME

B.1. Poor coverage of SME in curricula. Software maintenance has traditionally received very little attention from those involved in software engineering education as compared to its importance [7]. Also the later study of [8] in identified the fact that despite the needs, SME had still been generally overlooked in university curricula. More recently, the incorporation of SME in to the software engineering curricula related to the CCSE-initiative has been discussed in ICSM 2002. The then ongoing work of SEEK (Software Engineering Education Knowledge) included aspects such as: Configuration management, maintenance process models, reverse engineering, and impact analysis, their role as core of SME, proper time in curricula to introduce SME-issues etc.

Even today SME is rarely taught extensively in university-level before the third year of studies [9]. This is undoubtedly related to the fact that other issues have already reserved a relatively large portion of the curricula. This phenomenon has been identified also e.g. in 2001 in [5]. For example, in FIT there currently is not enough room for an SME-course in the curricula before the advanced-level of studies. Therefore, basic teaching of SME has to be organized otherwise. The SME-courses cited in this paper appear to have been only positive exceptions in this sense yet nowadays. In an ideal case students should be provided a versatile familiarization to the theoretical knowledge and status of the fresh scientific progress on the area sufficiently early in an interesting way.

B.2. Lack of educational materials. The second problem relates to the relative rarity of available and up-to-date educational materials. For example, experiences on teaching software reengineering at University of Leicester [10] included: Shortage of educational materials and difficulty of selecting suitable materials, tools, depth of coverage, and size of practical exercises. Students aiming at master's degree of computer science and nearby disciplines in university-level require both an up-to-date and versatile view of the field.

Although there are some good text-books, there is also a need to provide knowledge in an interactive way. There is no such single general text-book which would serve particularly well all these general needs. The SWEBOK Guide [2] provides a very relevant but short introduction. A relatively well-written and relatively new general-level treatise on the subject area is [14]. Management-oriented books include [1,12] but these books do not cover all the essential technical aspects of maintenance. There are also other relatively new books which however focus heavily either on practice; without proper description of the theoretical background, or on more general, loosely related aspects, such as software reuse [11].

B.3. Difficulties in producing realistic maintenance situations. The third and probably hardest problem is that realistic maintenance situations appear only in realistic environments. This problem has been attacked in many studies with various approaches. Experiences on teaching software maintenance related to a maintenance project of selected realistic software products at University of Western Ontario have been reported in [15]. The experiences included: Usefulness of control over group formation process, introducing central relevant concepts early enough in the courses, and interaction among the students beyond their own groups. A project-based course on teaching SME focusing on code comprehension methods has been organized at Oklahoma State University [16]. Also relations to industrial practice have been studied. A detailed process maturity model focusing on corrective maintenance for educating and training maintenance engineers has been developed at Stockholm University [17]. Finally, teaching SME by implementing change-requests on open-source systems has been experimented at Wayne State University providing realistic project experience [18].

Although practical project-based approach can be very valuable and desirable, as described e.g. in [15,16,18,19], dealing with real-world highly connected systems within the currently typical software engineering courses is often difficult in practice. This is due to the student's obviously limited knowledge and experience before third or fourth year of their studies, typical limitations of the available instructor resources, practical problems in organizing such endeavours etc. Even comprehending such real-world, large-scale legacy systems is an issue itself, and changing of those programs without negative side-effects is non-trivial to achieve and to evaluate. Therefore this problem is complex to be solved effectively.

Our experiences in this regard are similar to those presented in [10]. In practice, it appears to be hard or impossible to replicate the above-mentioned realistic maintenance situations in basic-level university courses. One central underlying reason to this is that the size of practical and applicative works, related to non-advanced level computer science courses typically is only about 1 cu. Within such time-frame only relatively simple changes of interfaces, and adding of functionalities could be performed. Therefore this problem can be effectively attacked only by organizing the applicative tasks related to larger dedicated work-projects, which are performed in an ideal case in cooperation with industrial partners [15] or related to open-source systems [18]. This approach has been followed also in our unit in general context by offering an option to all students majoring information systems science or computer science to attend to a large 10 cu. general work project during the third year of the studies.

C. Seminar-based approaches

As represented above, there are three main problems and two different kinds of issues to be taught; theoretical and applicative. Before the students can be expected to be mature enough to successfully commit themselves into the clearly needed large SME-projects they need to be provided with an introduction and to acquire some solid, theoretical science-based knowledge and understanding of SME. Therefore, in this paper we will focus on studying the feasibility and essential elements of providing that sort of knowledge within the above-described practical resource limitations often present in universities via our seminar-based approach.

Instead, the applicative aspects of SME are not our focus. Since there were no earlier reported studies on SME-seminars, we have organized seminars and systematically gathered empirical data concerning their characteristics.

III. SOFTWARE MAINTENANCE SEMINARS

Three software maintenance seminars have been organized. The general context of the seminars has been identical. The seminars have been organized as part of "Software Engineering" course in FIT. That mandatory course has mainly been targeted to students having computer science, information technology or similar disciplines as their major subject. The course has typically been taken during the second year of university-level studies. The instructor of all three seminars was the author of this paper. The substance area of each seminar has first been introduced in the course to the students via lectures by the seminar instructor. The actual seminar has followed that introduction. Seminar work tasks were then assigned to students to be performed as group works. They were mandatory to the students attending the course. Seminar groups organized themselves autonomously. The task of each seminar group was to analyze one scientific article, represent its analysis at the seminar, write a carefully finished summary report of the article, serve as opponent to one other group, and to listen the presentations of other groups.

A. General descriptive statistics

There were 127 groups, and 1-5 students in each group (Fig. 1). Total number of participants was 296. The basic statistics of the seminars (SMS-I, SMS-II and SMS-III) are shown in the Appendix, Table I.

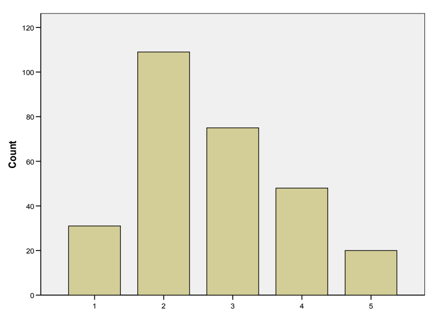


Fig. 1. Seminar group size (GSIZE).

A.1. Students. There were 282 students who proceeded to the actual seminars. 198 of these were men and 84 women. Average age was 24.3 years. Average year of studies at the university was 2.6 (median was 2). FIT grants degrees on information systems science (ISS), and information technology and computer science (IT). 93% of the students within the study were aiming at a degree within FIT. 82% of the students were focusing on information systems science.

The total accumulated amount of credit units for each student was determined, mean was 150 credit units, standard deviation = 47, N = 269. As noted earlier, students organized themselves into seminar groups. Median of group size was 3.

A.2. Articles. The seminars have covered most of the important issues in software maintenance field. There were three primary themes and eleven primary sub-themes. The students studied the articles as follows: Software maintenance techniques, including reverse engineering, reengineering, program slicing, and program visualization. Software maintenance tasks, inclusing software maintenance processes, program comprehension, maintainability improvements, and debugging. Economic estimation of software maintenance, including maintenance cost estimation, maintenance effort estimation, and software metrics and measurement.

There were 33 different publication series. All the studied papers have been published in good academic journals or conference proceedings. Complete bibliographic information of the articles offered to the students is provided in: http://www.cs.jyu.fi/~koskinen/sms-articles.htm

That set of articles has been selected by the seminar instructor based on his experience on the area. The Ph.D. thesis [20] included a large literature review. The over 400 references of that review have served as a starting point for the literature gatherings and selections of this study. The following selection criteria have been applied:

1) Language; all articles have to be written in English.
2) High scientific level; publication in refereed international scientific series.
3) Publication year; the articles should not be very old.
4) Relevance; all articles have to be on the area of SME.

Each individual student has been given an opportunity to freely select; from the predefined set, the paper which best matches his or her background, talents, and interests. However, since the article lengths vary, for the sake of fairness, there was one restriction: Larger groups were obliged to analyze longer articles than smaller groups. Therefore, n-person groups were offered only, in average, about n-times longer articles than 1-person groups. Also, for the sake of fairness, each group had to analyze a different article.

B. Learning objectives and evaluation

B.1. Learning objectives. The general nature of educational objectives can be classified e.g. based on so-called Bloom categories, which are described e.g in SWEBOK Guide [2, p. 1-7]. These categories concern increasing depth on: Knowledge, comprehension, application, analysis, synthesis, and evaluation. In our case, the students' general starting level has been such that they have had no earlier experience on software engineering related seminars or other SME-courses. For most of them this was also the first time they conducted academic group works and analyzed scientific articles written in non-native tongue. Therefore, the main general learning objectives have been to increase knowledge on SME, and capabilities on analyzing science-based articles of the field. The summary report has also included a portion prepared by the students evaluating each analyzed article.

The presentations have required the students to be able to summarize the main points of the analyzed method or technique. Implicit hopes and expectations have naturally included that the students would benefit from the presentations of the other groups in synthetic sense via increased and extended comprehension. On the other hand, since the analyzed articles have been scientific and the seminars did not intend to include applicative tasks, there were no applicative objectives. Since the tasks were performed mainly as group works communication and cooperation skills were implicitly needed during the process.

B.2. Evaluation. The instructor evaluated the groups works based on a standardized form (Appendix, Table II), which contains 15 criteria related to the report, and its presentation. We have planned that form in order to pay attention to all the central relevant aspects in these kinds of seminars during the whole lifecycle of the seminar works. The instructor provided feedback to the students multiple times during the lifespan of the seminar works. The assigned opponent group provided feedback at the seminar. During the first phase review of the seminar reports the instructor used the form as a basis for giving the groups further improvement suggestions. The instructor examined the report manuscripts two times.

The seminar instructor systematically monitored the seminar work quality based on the evaluation form. Also the students have been aware already prior to works of these criteria. Therefore, these criteria have been explicit operational learning objectives. The instructor controlled whether the students were listening to the presentations of other groups only in a sense of requiring participation to the seminar sessions. The instructor did the final evaluation of the report-related factors after the group returned its final version of the report. The instructor evaluated the seminar works group-wise.

C. Seminar points

Finally, the instructor considered the above-described criteria carefully, in detail, and aiming at impartiality for each group. The instructor used the evaluation form to support objectivity. Fig. 2 shows the results of the students.

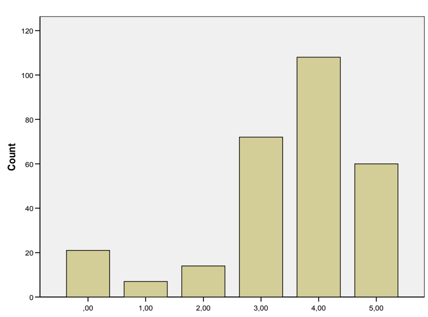


Fig. 2. Students' categorized seminar points.

The final result (SPS; Student's Points of Seminar) has been received as an average of the evaluated factors. This variable measures the success of individual students on the seminars based on the outcome of the group works. Fig. 2 shows the distribution of SPS compactly as categorized (rounded downwards to integers) into six classes.

The applied scaling roughly corresponds to international grading systems: A ... F: A = 5 = excellent, B = 4 = very good, C = 3 = good, D = 2 = satisfactory, E = 1 = sufficient, F = 0 = fail. Other SPS characteristics were, mean = 3.55, minimum = 0, maximum = 5.30, standard deviation = 1.33, N = 282. 107 groups completed their works. 20 groups could not follow the schedule. The seminars produced 683 pages of thoroughly inspected and published seminar reports.

IV. Discussion

This section discusses and explains the objectives of the study (*A*), and general results and experiences (*B*), regarding the general problems of teaching SME (*B.1*), teaching SME-specific issues (*B.2*), organization of the seminars (*B.3*), and relation to other studies and research (*B.4*).

A. Objectives

We have focused on studying the general feasibilility of SME-seminars. We have gathered novel experiences of the seminar-approach with a relatively large number of students. Seminars have been characterized by the scientific nature of the analyzed articles, extensive coverage of the central SME-aspects, and interactivity. Instead, the study has purposefully not included technical application of maintenance skills. That sort of activities are not convenient to be handled in these sort of seminars. Those activities have been studied in other unrelated studies, whereas there are no earlier systematic studies on SME-seminars. The general nature of the main learning objectives has been to increase knowledge on SME and skills on analyzing scientific articles. The students' task consisted of five sub-tasks. Due to the modest starting level of the students the set objectives were not trivial for them.

B. General results and experiences

B.1. General problems of teaching SME. There are three general problems of teaching SME as identified in Section II. The general experiences from this study show that seminars can especially alleviate the problem of lack of dedicated educational materials (problem *II.B.2*). We have shown that seminars effectively enable covering a large spectrum of SME-issues, including recent scientific innovations. They also enable interactivity, which is generally not supported by educational materials as such. Interactivity is important since it helps to increase students' interest to the area. Number of the active students increased in the subsequent seminars showing interest towards the seminars. Presentations of the groups activated well the "Software Engineering" course. Seminar works also increased collaboration among the students which is relevant since many of the SME-tasks require collaboration.

The experimental results concerning seminar success help to pay attention to the relevant issues while planning similar seminars. Knowing those issues encourages teaching SME more extensively and enables to teach it more successfully in university-level (problem *II.B.1*) using interactive forms of education. Seminars also help to introduce SME since it is generally hard to find enough room for large SME-courses in the non-advanced level of the curricula. The problem of reproducing realistic maintenance situations (problem *II.B.3*) in turn has not been attacked in these seminars. As mentioned earlier, that sort of applicative activities should be provided in separate and relatively large advanced-level work projects which are much more suitable to teaching such issues.

B.2. Teaching SME-specific issues. The covered SME-specific issues were listed in Section *III.A*. The students need to have solid science-based knowledge on these issues, which the seminars provided. There were three general themes identified: a) Software maintenance tasks, b) software maintenance techniques, and c) economic estimation of software mainte-nance. a) Concerns processes, b) relates to process support mechanisms, and c) relates to evaluation of the processes. These are generic issues typical also to other kinds of software engineering activities. Therefore, it is likely that the maintenance-specific issues and other software engineering issues; such as software development, do not differ significantly in a sense that they should be taught in a significantly different way. On the other hand, it is likely that interactivity as such will be needed also in the general software engineering field. Therefore, the interactive nature of the seminars should be useful also in that more general context.

It is very likely that the applied seminar format is suitable also to dealing with software engineering issues in general with similar kinds of general objectives. This is due to the fact that the seminars were based on scientific articles and the applied high-quality scientific format of representing results in them is refined, flexible and capable of representing probably almost any kind of theoretical issues in a relatively communicative way, apparently also from the view-point of the second year students. On the other hand, it is clear that teaching SME as such requires also other kinds of forms of teaching, including the already discussed larger work-projects.

B.3. Organization of the seminars. Seminars enable much content and activity within the typically limited amount of available space in the curriculum. However, they also require relatively large amount of instructor resources. Based on our experiences, the practical maximum of groups per term for a single qualified instructor to deal with within the typical resource limitations is around 30-50. Controlling 45-50 groups effectively requires in practice that the instructor has deep knowledge on SME; Ph.D. or equivalent, and earlier experience on similar seminars. While dealing in this case with 45 groups, the required instructor resources are about 140 h. Seminar success in this level, including analysis of scientific papers, especially while written in non-native language, requires detailed instructions.

Preparation of each seminar took about 15% of the total time. Seminar sessions as such took about 10%. Therefore, the seminar sessions as such were the high-lights of the relatively extensive background effort. More time to the sessions would be clearly beneficial if it could be arranged. Increasing the relative duration of the sessions is quite possible if the level of expectations concerning the quality of the reports is reduced in turn. In these seminars the expectations were relatively high in that regard. Two-pass reviewing and commenting of the produced reports took about 65% of the total time. General communication and coordination took the remaining 10%. These resource requirements need to be heeded while organizing similar kinds of seminars.

About 84% of the groups passed the seminar. Therefore, the seminars were successful. About 90% of the failures were related to small (1-2 person) groups. Therefore, group formation is a very important issue. Communication is needed in these kinds of seminars and isolation of the students in this sense should be avoided. Special attention should be paid on generally not allowing singleton groups and by monitoring the time-table of the group formation process.

B.4. Relation to other studies and research. Bringing the scientific perspective forth already at this stage of the studies supports the writing of master's theses. These seminar works have in many cases already served as a starting point for theses. This is generally especially hoped for, since because of well-known economical, and other reasons, students typically have problems in completing their theses. Timely completion of studies in turn is a central consideration related to the public funding of many universities. Since there were 127 prominent articles the students received a very good basis for studying the literature of the area further in the theoretical parts of their oncoming theses.

Free selection of the articles has in its part enabled focusing on the likely areas of the oncoming theses. It is also a generally hoped goal to bring teaching and research together. Seminars have been organized around a main theme. Seminars have in part related to the research projects of FIT, namely: HyperSoft (Hypertextual Software Maintenance), and ELTIS (Extending the Life-Time of Information Systems). The archived and published seminar reports enable long-term improvement of teaching SME within FIT. The results of the ELTIS-project have already been transferred as a part of teaching related to SME in FIT recently.

V. CONCLUSIONS

SME is a wide, economically important, and nowadays also rapidly advancing sub-field of software engineering. University-level computer science students should be given versatile and up-to-date science-based SME-knowledge sufficiently early during their studies in an interactive and motivating way. Reaching these ideals based on the currently available educational materials and within the typical size-limitations of curricula is hard. Seminar-based approaches clearly support these goals well, but there have been no earlier systematic studies on SME-seminars. Therefore, there has been a clear need to study the feasibility of SME-seminars.

We have organized SME-seminars and systematically gathered experiences. The seminars as such were successful and clearly verified the general feasibility of the seminar approach to handle these kinds of issues. Seminars and the group works include interaction which supports keeping up students' motivation, which in turn is important in this area. Motivation was supported also by allowing the students to focus on the sub-areas which interested them most.

The seminars have helped to create links between general teaching, writing of academic theses, and ongoing research which are important for further improvements of the university-level teaching in this area. The general success of the seminars hopefully encourages others to study teaching of SME based on similar approaches. Taking the represented experiences and lessons into account supports planning and organizing feasible seminars in software engineering.

References

[1] R. Seacord, D. Plakosh, and G. Lewis, *Modernizing Legacy Systems*. Addison-Wesley, 2003.

[2] A. Abran, J.W. Moore, P. Bourque, and R. Dupuis (eds.), *SWEBOK: Guide to the Software Engineering Body of Knowledge 2004 Version*. IEEE Computer Society, 2004.

[3] *Computing Curricula 2001 Computer Science: Final Report (15 December 2001)*. IEEE Computer Society and Association for Computing Machinery, 2001. http://www.computer.org/portal/cms_docs_ieeecs/ieeecs/education/cc2001/cc2001.pdf

[4] *Software Engineering 2004: Curriculum Guidelines for Undergraduate Degree Programs in Software Engineering: A Volume of the Computing Curricula Series (23 August 2004)*, IEEE Computer Society and Association for Computing Machinery, 2004. http://sites.computer.org/ccse/SE2004Volume.pdf

[5] J.P. Myers Jr, "Software engineering throughout a traditional computer science curriculum", *Journal of Computing Sciences in Colleges*, vol. 16, number 2, pp. 31-41, 2001.

[6] B. Bracken, "Progressing from student to professional: the importance and challenges of teaching software engineering", *Journal of Computing Sciences in Colleges*, vol. 19, number 2, 358-368, 2003.

[7] B.J. Cornelius, M. Munro, D.J. Robson, "An approach to software maintenance education", *Software Engineering Journal*, vol. 4, number 4, pp. 233-236, 1989.

[8] M. Newby, "Legacy systems, software maintenance and computing curricula", *Proc. Software Education Conf.* IEEE, 1994, pp. 96-102.

[9] *Ohjelmistotekniikan Korkeakoulutuspäivä* (in Finnish). Tampere, Finland: University of Tampere, 2002.

[10] M. El-Ramly, "Experience in teaching a software reengineering course", *Proc. 28th Int. Conf. Software Engineering* (ICSE 2006). ACM, 2006, pp. 699-702.

[11] S. Jarzabek, *Effective Maintenance and Evolution: A Reuse-Based Approach*. Auerbach, 2007.

[12] T.M. Pigoski, *Practical Software Maintenance: Best Practices for Managing Your Software Investment*. Wiley, 1996.

[13] M. Polo, M. Piattini, and F. Ruiz, *Advances in Software Maintenance Management: Technologies and Solutions*. IGI Global, 2002.

[14] P. Grubb, and A. Takang, *Software Maintenance: Concepts and Practice* (2nd ed.). World Scientific, 2003.

[15] J.H. Andrews, and H.L. Lutfiyya, "Experiences with a software maintenance project course", *IEEE Transactions on Education*, vol. 43, number 4, pp. 383-388, 2000.

[16] M.A. Austin III, and M.H. Samadzadeh, "Software comprehension/ maintenance: An introductory course", *Proc. of the 18th Int. Conf. on Systems Engin.* (ICSEng 2005). IEEE, 2005, pp. 414-419.

[17] M. Kajko-Mattsson, S. Forssander, and U. Olsson, "Corrective maintenance maturity model (CM^3): maintainer's education and training", *Proc. 23rd Int. Conf. Software Engineering* (ICSE 2001). ACM, 2001, pp. 610-619.

[18] J. Buchta, M. Petrenko, D. Poshyvanyk, and V. Rajlich, "Teaching evolution of open-source projects in software engineering courses", *Proc. of the 22nd IEEE Int. Conf. on Software Maintenance* (ICSM 2006). IEEE, 2006, pp. 136-144.

[19] K.R. Pierce, "Teaching software engineering principles using maintenance-based projects", *Tenth Conf. on Software Engineering Education and Training* (CSEE&T 1997). IEEE, 1997, pp. 53-60.

[20] J. Koskinen, *Automated Transient Hypertext Support for Software Maintenance*. *Jyväskylä Studies in Computing*, vol. 4. Jyväskylä, Finland: University of Jyväskylä, 2000.

Appendix: Summary of the Seminars

Table I. Software Maintenance Seminars

Aspect	SMS-I	SMS-II	SMS-III	Total
Involved groups	28	45	54	127
Active groups	24	37	46	107
Students	76	111	109	296
Seminars	10 h.	13 h.	15 h.	38 h.
Organization	75 h.	105 h.	125 h.	305 h.
Total used teaching resources	85 h.	118 h.	140 h.	343 h.
Reports produced	173 p.	251 p.	259 p.	683 p.

Table II. Evaluation Form

Number	Criteria	Details	Criteria group
1	Layout of the report	There were strict instructions for the coherent format of the reports.	Report
2	Extent of the report	The size of the report within the allowed range specified for the article.	Report
3	Relevance of the report's focus	The report had to focus on the issues inside the article.	Report
4	Quality of language	Use of complete sentences, readability, style etc.	Report
5	Correctness of the details	Accuracy of the presented assertions, numerical data, references.	Report
6	Definition of the central concepts	Their presence and correspondence to the original ones.	Report
7	Description of the research problem	The problem as represented in the article.	Report
8	Description of the hypotheses	For empirical works.	Report
9	Description of the applied research methods	Procedure of the study.	Report
10	Description of the results and their significance	Main results and their implications.	Report
11	Group’s own observations and conclusions	Students' evaluation of the article.	Presentation
12	Clarity of the presentation	Presence of the main points, oral and audio-visual appearance.	Presentation
13	Interaction between the group and audience	Answers to the opponents and others.	Presentation
14	Conformance to the predefined schedules	Meeting the mile-stones for reviews etc.	Presentation
15	Action of the group in the role of opponents	Constructive comments.	Presentation

A Vector Based Smoothing Algorithm of High Speed Motion at Trajectory Corner

Lei Luo, Jun Hu
School of Mechanical Engineering, Shanghai Jiaotong University, Shanghai 200240, China
e-mail:luolei@sjtu.edu.cn

***Abstract*- To ensure the quality of products and efficiency of machining and to protect the machine tools from damage in high speed working conditions, one of the key solutions is to smooth the feedrate profile at the corner of two adjoining moves. In the paper, a vector based look-ahead algorithm is presented for feedrate smoothing at trajectory corner by interpolating adjoining moves simultaneously. By analyzing and simulating, it is verified that this algorithm can realize good machining quality and smooth velocity at the corner of the tool path with high speed.**

***Keywords*- Feedrate Smoothing, Vector Based Transition Curve, High Speed Machining**

I. INTRODUCTION

CNC machining with high speed and high accuracy is the key solution to assure the machining quality and efficiency [1, 2, and 3]. To protect the CNC feed drive system from great impact load in high speed machining condition and avoid quality problem of the work piece, the feedrate smoothness of the CNC machine tools should be guaranteed [4]. When the corner between two adjoining moves is rough, the variation of the velocity vectors will become larger and the great impact will be generated in high speed condition. In order to solve this problem, many CNC systems decelerate until the feedrate reaches zero at the end point of the corner and then continue to move with acceleration.

However, the frequent deceleration and acceleration reduce the machining efficiency severely. In order to generate smooth feedrate profile of the trajectory, multi segment feedrate smoothing treatment instead of single segment feedrate profile generation is required.

Figure 1 illustrates the single segment feedrate profile with acceleration, uniform velocity and deceleration stages. Figure 2 illustrates the feedrate profile with multi segment feedrate smoothing treatment. The adaptive algorithm to control the velocity at the corner should be proposed to meet the demands of impact load restriction and machining efficiency requirement. Hu et al. [5] proposed a look-ahead algorithm to control the velocity at the corner, which calculates the angle of the corner in advance and adjusts the velocity correspondingly to restrict the shock of the machine tools caused by variation of the velocity vectors within the accepted ranges. The algorithm can avoid any stop at corners on the tool path so that good motion performance can be achieved and high working efficiency can be obtained. But its calculation is complicated and the requirements of the hardware are high. Williams [6] developed a vector blending method to remove instantaneous velocity changes during abrupt direction changes. Zhang et al. [7] developed a method to handle the problem by adding an extra segment of arc at the corner to avoid stopping at the intersection points. This method avoids the frequent deceleration and acceleration in some degree, improves the machining efficiency and reduces the shock. But the maximal feedrate is also limited by the radius of the arc and the contour error is larger at rough corner.

In this paper, a vector based smoothing algorithm is presented for trapezoidal velocity profile, which can avoid over-cutting at corners effectively. The contour error can be restricted to a low level in high speed machining condition, while efficiency and surface quality of work piece can be guaranteed. In section 1, the vector based transition curve generation method is presented. In section 2, the maximum allowable feedrate and the maximum contours error of the transition curve are analyzed. Simulation results and conclusions are given in section 3 and 4 respectively.

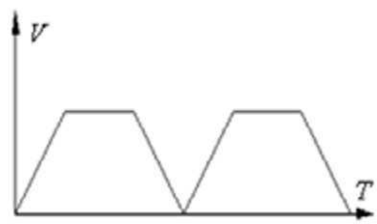


Fig. 1 Single Segment Feedrate Profile

Fig. 2 Multi Segment Feedrate Profile with Feedrate Smoothing Treatment

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_63,

II. Generation of Vector Based Transition Curve

According to the theory of space vector, the vector based transition curve can be generated by interpolating adjoining moves simultaneously. As the result, the smooth feedrate on the track at corners can be realized.

Assuming the directions of the adjoining moves $S1$ and $S2$ are shown in figure3, the transition curve is begun to be generated when the feedrate of $S1$ is equal or smaller than the maximum allowable velocity V_{blend} at the corner. In transition curve generating period, the interpolating feedrate of $S1$ decreases from initial velocity of V_{blend} to zero, while the one of $S2$ increases from zero to V_{blend}. During every interpolating period, assuming the displacement of $S1$ is $\overrightarrow{m_i}$ and the one of $S2$ is $\overrightarrow{p_i}$, the new displacement vector $\vec{l_i}$ can be generated by adding the two displacement vectors, i.e.

$$\vec{\boldsymbol{l}}_i = \overrightarrow{\boldsymbol{m}}_i + \vec{\boldsymbol{p}}_i \tag{1}$$

The transition curve generated according to the above method not only makes track and feedrate smooth, but also guarantees the contour error near corners within the acceptable level and assures high tracking accuracy.

The transition curve between linear and circular trajectory is

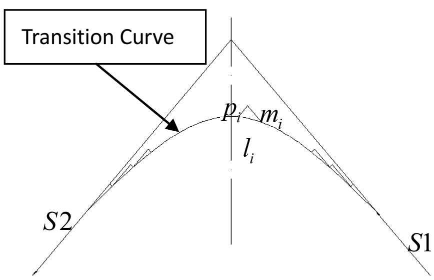


Fig. 3 Transition Curve Between two Linear Trajectories

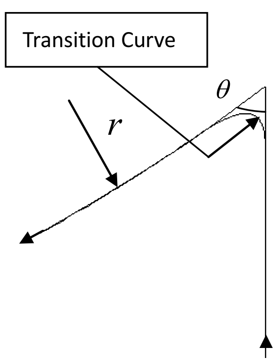


Fig. 4 Transition Curve between Linear and Circular Trajectory

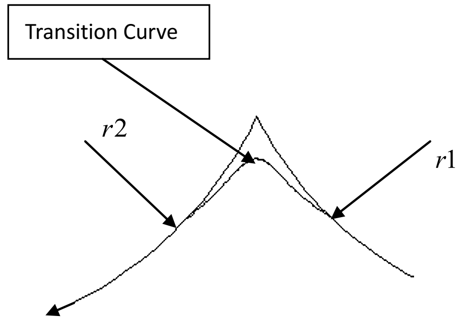


Fig. 5 Transition Curve between Two Circular Trajectories

shown in figure 4 while the transition curve between two circular trajectories is shown in figure 5.

According to the analysis above, the generated displacement vectors in the transition process at corners during every interpolating period ΔT is $\vec{l_i}$. So the velocity can be calculated as follow,

$$\overrightarrow{\boldsymbol{v}_i} = \frac{\vec{\boldsymbol{l}}_i}{\Delta \boldsymbol{T}} \tag{2}$$

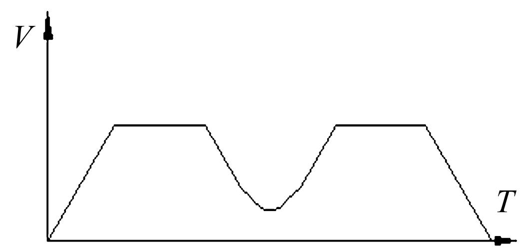


Fig. 6 Feedrate Profile of Transition Curve

In the transition process at corners, interpolations of adjoining moves are proceed simultaneously so that the feedrate will not decrease to zero and the frequent deceleration and acceleration can be avoided. Suppose each segmental trajectory is continuous, then the generated transition curve is continuous and differentiable and the feedrate at corner is continuous and smooth. Assuming feedrate of every segmental trajectory follows trapezoidal velocity profile, the velocity curve at corners is illustrated in figure 6.

III. Transition Curve Maximum Feedrate Determination and Contour Error Analysis

The vector based transition curve algorithm takes the feedrate smoothing and contour error controlling in consideration. The maximum demand distance for feedrate deceleration can be calculated in advance depending on the contour error at the corner. Then the maximum allowable feedrate at the beginning of the transition curve can be obtained. In figure 7, assuming every movement is based on trapezoidal velocity profile, σ is the angle of the adjoining moves vectors, $\vec{v_1}$ is the unit vector at end point of $S1$

and $\overrightarrow{v_2}$ is the unit vector at start point of $S2$. Therefore,

$$\sigma = \overrightarrow{v_1} \cdot \overrightarrow{v_2} \tag{3}$$

Then we can get,

$$2\theta = \pi - \sigma \tag{4}$$

Between liner and circular trajectory or between two circular trajectories, the unit vector at starting or ending point of the arc is the tangent vector of that point.

When feedrate at corners is remained constant and σ is large, the contour error is large and the shock of the machine tools is great. It will affect the quality of work piece and may do harm to the machine system. If large feedrate variation exists, the tracking and contour accuracy cannot be guaranteed within the high level according to the effect of inertia force and the working efficiency of the machine tools cannot be high. To avoid above problems, a new algorithm will be discussed in the following text.

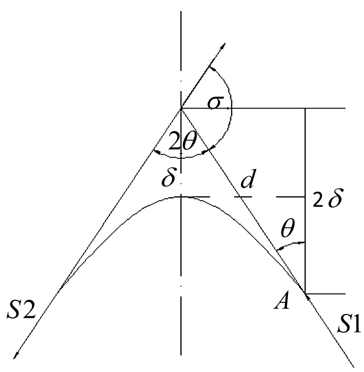


Fig. 7 Contour Error at the Corner

It is assumed that the feedrate of $S1$ decreases to zero after passing through the moving distance of d, which is the length of the line or the arc, and it is also assumed that a is the acceleration of the trapezoidal velocity profile. The shortest distance from the intersection point of the corner to the transition curve is δ. Then d can be calculated as follows,

$$d = \frac{2\delta}{\cos\theta}. \tag{5}$$

According to the assumption above, the velocity at A can be calculated as follows,

$$V = at. \tag{6}$$

Therefore,

$$t = \frac{V}{a}. \tag{7}$$

Then, we can have

$$d = \frac{1}{2}a\left(\frac{V}{a}\right)^2 \tag{8}$$

Therefore V_{blend}, which is the beginning velocity of the transition curve at corner, can be calculated by (5), (6), (7) and (8).

$$V_{blend} = 2\sqrt{\frac{a\delta}{\cos\theta}} \tag{9}$$

From the above formula, it is clear that δ is the direct ratio to V_{blend} and V_{blend} can be restricted within limits to realize the needed contours error δ.

IV. SIMULATION

To verify the algorithm above, simulation is presented in this section. It is assumed that the programmed velocity, the maximum contours error, the maximum acceleration and the interpolation sample time are $V = 400mm/min$, $\delta = 0.5mm$, $a_{\max} = 100mm/s^2$ and $\Delta T = 2ms$ respectively. The data of tool path to be followed in XY plane are listed in Table 1.

Following the linear trajectories in table 1 with stop at every corner, the track represented as thick line is illustrated in figure 8 and the corresponding velocity profile is illustrated in figure 9. Moving along the same trajectories using the smoothing algorithm discussed above, the track represented as thin line is illustrated in figure 8 and the corresponding velocity profile is illustrated in figure 10.

TABLE I
The Straight Lines to be followed (in Absolute Coordinate)

X(mm)	0	1	2	3.5	5.5	8	11	14.5	18.5	23	28
Y(mm)	0	5	0	5	0	5	0	5	0	5	0

By analyzing the velocity profiles we can have following conclusions.

1) The productivity by using the vector based smoothing algorithm is increased about 11% comparing to the one without using the algorithm.
2) The angle of the corner is smaller, the response of the velocity at corner is better and the productivity is higher.
3) The transition curve generated according to the method presented above is closed to the actual contour at the corner and the contour error is small.
4) The shock of the machine tool at corners is reduced because of the smoothness of the track and the velocity. This improves the machining quality and protects the feed system of machine tools from damage.

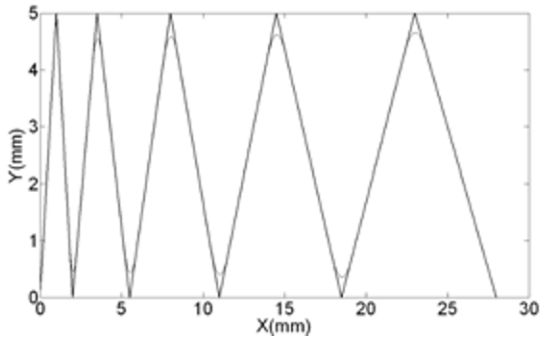


Fig. 8 Tracks with and without Using Smoothing Algorithm

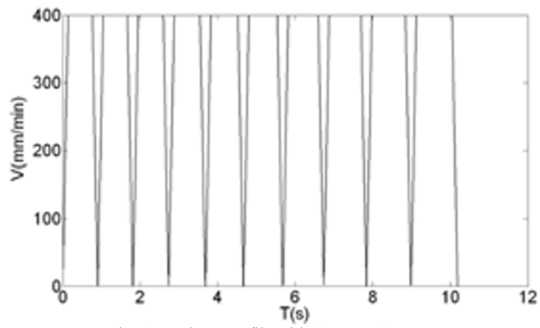


Fig. 9 Feedrate Profile with Stops at Corners

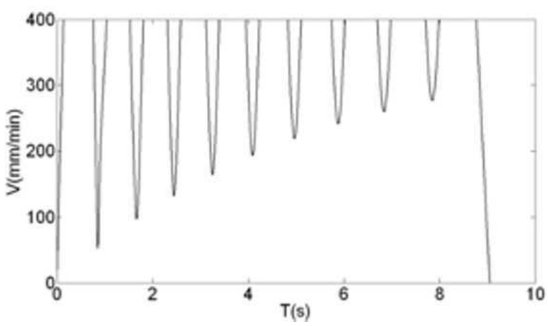


Fig. 10 Feedrate Profile by Using Smoothing Algorithm

V. CONCLUSION

Making feedrate smooth along the trajectory at the corner of two adjoining moves is the key technique to protect the machine tool from damage and assures the machining quality and efficiency in high speed machining condition. The vector based smoothing algorithm presented in this paper can achieve good performance at corners, whose advantages are easy to realize, smooth of the transition curve, easy to control the contours error at corners，small shock of the machine tools and good velocity response. The algorithm can fit the high speed machining well and it has been applied to the newest embedded CNC system in practice.

ACKNOWLEDGMENT

This project is sponsored by The Shanghai Science and Technology Development Foundation (NO. 05DZ22003) and (NO. 06DZ11418).

REFERENCES

[1] Z.B. Tang, X.Q. Tang, and B. Li, “Research and Implementation of High Speed and High Accuracy Motion Control Method in CNC”, Manufacturing Technology and Machine Tools, March 2001, vol. 3, pp. 32-35

[2] Y. Koren, “Computer Control of Manufacturing Systems”, McGraw-Hill, New York, 1983

[3] D.C. Wang, C.X. Li. “Research on Application and Intrinsic Fact of Self-Adaptive Control of CNC ”, Die and Mould Technology, 2001, vol. 6, pp. 69-72

[4] K. Erkorkmaz, Y. Altinatas, “High Speed CNC System Design. Part 1: Jerk Limited Trajectory Generation and Quintic Spline Interpolation”, International Journal of Machine Tools and Manufacture, July 2001, vol.41, no.9, pp.1323-1345

[5] J. Hu, L.J. Xiao, Y.H. Wang, Z.Y. Wu, “An Optimal Feedrate Model And Solution Algorithm For High-Speed Machine Of Small Line Blocks with Look-Ahead”, International Journal of Advanced Manufacturing Technology, 2006, vol.28, no. 9, pp. 930-935

[6] Chris Williams, “Vector Blending Cuts Cycle Times-Big Time”, March 2000, vol.72, no.9, pp. 192

[7] D.L. Zhang, L.S. Zhou, “Adaptive Algorithm for federate Smoothing of High Speed Machining”, Acta Aeronautica Et Astronautica Sinica, January, 2006, vol. 27, no. 1, pp. 125-130

Teamwork and ABET Review: A Template for Assessment

Karl Perusich, Ph.D.[1], Beverley Davis[2], Kevin Taylor[3]

Abstract -The latest methodology for ABET accreditation of engineering and technology programs requires the assessment of, among other things, teamwork skills. However, determining and assessing the skills of the team members is problematic. To be effective, engineers and technologists must work in teams to reduce the knowledge load, to share the work so that it may be completed in a timely fashion and to provide a means for error checking. Each of these reasons requires a particular skill set. It is natural to assess for an ABET review the extent to which these skill sets are being taught in a curricula. But what skill sets should be assessed. In this paper critical skills that differentiate working in teams from working in groups are identified. These skill sets can then be used as a template in constructing an assessment methodology for "teamwork".

Keywords: Accreditation, Team Work

I. INTRODUCTION

Engineering and engineering technology programs are required as part of their ABET assessment to demonstrate that their students have developed the requisite skills to function in a team environment. Although each program individually can set their standard and methods for making this determination there still should be a common model used. Unfortunately this model fails to distinguish between groups and teams, and any assessment made will miss the mark. In some sense the best approach to use in setting the standard is to determine why engineers and technologists need to work in teams. From this a skill set can then be enumerated. The skill set will then define topics, techniques and other artifacts that students should master and, as such, should be taught in classes they take. The mastery of these topics, techniques and other artifacts by the students then provides the basis for critical assessment of "teamwork skills".

Although seemingly straightforward an important consideration when following this top down approach to setting the assessment standard for teamwork is deciding whether the skill can in fact be taught in a traditional lecture/lab format used in most technical classes. If it can not, then the efficacy of the assessment method must be questioned. Technical educators tend to view the educational environment with blinders on assuming that, like equations or software, teamwork can be mastered via example and analysis. Unfortunately this tends not to be the case. The structure of teams and the aspects that make them successful, defined below, are not normally or easily reproduced in the classroom context in which teamwork is being taught.[1]

II. THE NEED FOR TEAMWORK

Engineering, unlike many professions, almost by definition must be conducted in teams. Engineers and technologists provide solutions at various levels of a system chain. Individual devices get embedded in systems that get embedded in large systems and so on. The solution developed generally has a time constraint that provides a temporal boundary on its applicability. The engineer has a narrow window in which a solution is viable and feasible either because the problem passes or it changes its nature. These two characteristics, large system solutions and timeliness of their delivery make working in teams an imperative for engineers for three reasons: breadth of expertise required, breadth of workload needed and error checking.

Engineering solutions generally are embedded in larger systems. These larger systems require the knowledge and expertise of a number of disciplines and fields. For example the design and construction of a modern passenger aircraft has applications for almost every recognized field of engineering. No person could ever be expected to have the cognitive capability or experience to have adequate knowledge in all of these different areas. The only effective way to manage the knowledge requirements is to distribute it over a number of individuals organized as teams.

Just as teams reduce the intellectual load for individuals, using this organization also reduces the absolute workload for engineers. As stated previously engineering is a time-critical enterprise. The technical solution developed has a narrow "window of opportunity" in which it is effective and applicable. Even if a single individual had the education, background and experience to produce an adequate solution to a problem, they may take too long to

1 Purdue College of Technology-South Bend, South Bend, IN, 46634. Perusich@purdue.edu
2 Purdue College of Technology-Lafayette, Lafayette, IN 47903. bevjd@purdue.edu
3 Purdue College of Technology-Kokomo, Kokomo, IN 46904. kdtaylor@purdue.edu

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_64, © Springer Science+Business Media B.V. 2010

complete it for it to be effective within the "window of opportunity" presented by the problem. Working in teams speeds the process by taking advantage of synergies among team members and between teams and by conducting activities in parallel wherever possible.

Another important but often overlooked reason for working in teams is peer review and error checking. Engineered systems have the potential for disaster if they are inadequate putting in jeopardy potentially many lives. Working in teams provides a natural (and non-threatening) means of error checking and peer review. Individual solutions will be presented to the team or other teams for critical evaluation as part of a normal process of design review. This process, inherent in teamwork in engineering, will provide critical feedback about the solution and the process used to achieve it. By working in a multi-disciplinary environment, this feedback comes from not only areas of the engineer's technical expertise but also from a broader technical viewpoint provided by other teams. In most cases this then eliminates potential flaws in the design.

III. TEAMS VS. GROUPS

Students often participate in projects or work together in groups on a task or design in a technical class. But working in groups is not the same as working as a team. Teams and groups both bring together multiple individuals to accomplish a task, but have little else in common. Teams develop the bonds and framework within which the members participate well beyond the expectations of a group. Since many academic assessments involve examining the mastery of skills of individuals working in groups rather than teams, it is worthwhile to examine the differences in detail.

A group is a number of individuals put together (voluntarily or involuntarily) to accomplish a goal.[2] Goals for a group will be discussed in general terms, but generally lack depth and clarity. Groups tend to be a short expediency for accomplishing a task so members can and will lack a history and experience of working together. Lacking with groups is both the expectation and realization of synergy among the group members in accomplishing a task that comes with a common history and experience. [3] With groups the sum tends to better no better than the parts. The reward structure for groups tends to remain individualistic. Rewards and penalties are given for individual contributions to the group effort. Individuals within the group feel success if their contribution was successful even if the overall group accomplishment was not. Finally, individuals in a group tend to fail to collaborate and coordinate their activities. Individuals could shuffle in and out of a group without necessarily degrading its performance.[2]

Teams, on the other hand, develop beyond the individual capabilities and contributions of its members. The sum is really greater than the whole for a team. Shuffling individuals in and out of a team would seriously degrade its performance. Much of this lies in the key differences between teams and groups. Teams have clear, concise goals that drive all aspects of their activities.[5] These goals can have two dimensions. One is the external requirement set by the task to be accomplished. The second dimension is an "internal" set of goals associated with the shared psyche of the team members; the idea of "taking one for the good of the team". Unlike groups self-sacrifice for the good of the team is normally an expected behavior.

Key to effective teamwork is the synergy that comes from the members in the team. Through the history of working together team members develop an instinct about the capabilities and expectations of other team members. They will undertake tasks with the expectation that other team members will support their efforts without being asked or instructed. One team member will mask or make up the deficiencies of another. Individuals clearly understand what they are to do and, as important, clearly understand what other team members will do. Members work together producing a "team effort", i.e. a synergy of efforts and accomplishments not present if they worked individually.

The reward structure for a team is significantly different from that of a group.[6] The external environment in which the team is embedded, more often than not a larger team, holds the team accountable for its actions and accomplishments, not individual members. The "team" is rewarded when successful and the "team" is punished when its fails. Not only is the team externally held accountable, but internally a culture will develop that provides for mutual accountability. Individual team members will have the expectation that other members will work for the "good of the team" and that they hold each other mutually accountable for accomplishing this.

Roles and responsibilities, whether formally or informally, are clearly defined in a team. This contributes to the synergy that makes a team stronger than its individual elements. Individual roles are generally defined in a team in such a way that capitalizes on the individual's strengths while minimizing their weaknesses. Individuals in a team will work on those aspects of a task where they conceivably can make the greatest contribution to the team effort, because of their training, expertise or experience. This produces an overall maximization in the team effort because each aspect of accomplishing the goal is being handled by the individual that has the greatest chance to succeed at it.

Since groups are likely to be transitory in nature, they are organized informally with few rules. Any boundary conditions they operate with will develop as an expediency to complete the assigned task. Teams, unlike groups, have an anticipation of lengthy collaboration so rules and regulations are more systematically developed, formalized and communicated. [7] This helps reinforce the individual and inter-individual understandings of the role they each play in the functioning of the team, further optimizing its overall performance.[4]

Finally, although members in a group will out of necessity interact, each will tend to see and focus on only their contribution. Teams, not only interact, they also collaborate and coordinate. Members of a team understand how their actions within the team environment affect and are affected by the actions of the other members. Unlike groups, teams will discuss not only the division of the task into manageable parts, they will also discuss the organization of the team itself. Groups tend to focus only on the task. Teams will focus as much on their organization as on accomplishing the task. [7]

IV. TEAMS

The key to understanding how these differences between teams and groups come about is recognizing the transitory nature of groups and the anticipatory collaboration of teams. Groups are put together, many times on an adhoc basis, to accomplish a task. The membership might be chosen to cover certain skills and capabilities recognized as necessary for meeting the goals set out for it. There is no expectation that its members have worked together before and there is no anticipation that the group will survive as an entity after the task has been accomplished.

Teams, individually and collectively, anticipate surviving beyond a short-term goal or task. Individuals are admitted to a team as much on whether other members think they will contribute as on the skills that they bring. [4]This is because there is an anticipation that teams will be long lived so there will be a lengthy period when their individual members must interact. The effectiveness of the team is as much about how its members understand roles, the individual members accepting the responsibilities and accountabilities of their individual roles, and their interacting collectively with a commitment to achieve with the skills the team has.

Because teams are long-lived they will share a history of interaction in which they learn to collaborate. During this history each team member will learn their limitations within the framework of the team, the limitations of other members, and how their different limitations can be compensated for. This mutual understanding contributes to the effectiveness of the team being greater than the capabilities of each member individually. In fact, this group learning may be explicitly made part of the team building process through practice. For example, sports teams generally have long periods between tasks, i.e. wining over their competition, during which they practice together to refine their roles in the overall functioning of it, i.e. the game plan. Without this shared history, which can only be developed with time, many of the attributes listed above would be impossible.

V. TEACHING TEAMS

The attributes outlined above can certainly be presented in a traditional lecture/lab classroom setting. But the *development* of teams in a classroom is more problematic. Two constraints could preclude creating a practical experience for students in teamwork in the same way that technical educators can give students "hands on" experience in electrical circuits or fluid power. First, many of these attributes require historical experience by team members of working together, if not practice in accomplishing the goal. Second, the reward structure required for evaluation in the classroom is significantly different from the reward structure for team evaluation. University students receive a grade for their class, so contributions must be judged and evaluated individually. Teams, though, will be rewarded for the collective effort.

The semester structure used in most university settings requires materials be presented, absorbed and evaluated 2-3 times weekly for 14-16 weeks. The exact student makeup of each class will vary from semester to semester. Also, students will take multiple classes, each potentially having different members. Thus, students will be constantly shifting their attention and resources throughout the week to different topics and interact with different individuals. True team development requires significant interaction, even practice. The same individuals must work together on the same task, otherwise learning each others individual limitations becomes impossible. The fact that students must take multiple classes in a limited time frame makes this problematic.

Students taking university classes must receive a grade for their effort. This judgment is based on their individual contribution or individual mastery of the materials in the course. Teams are judged by their collective effort. The *team* is evaluated, not its individual members. Success or failure is judged at the team level. This in some sense produces a dichotomy that clouds the teaching of teams. Teamwork must be assessed at a collective level, but students must be graded at an individual level.

VI. TEMPLATE FOR TEAMWORK ASSESSMENT

Ideally those skills necessary to work in the demanding team environments of today's global economy would be taught in the classroom. These skills could then be seamlessly evaluated for an ABET review. But the inherent limitations of teaching these skills in a traditional classroom environment provide obstacles to an effective accreditation assessments. Any effective evaluation technique must recognize these boundary conditions and work within them. With these boundary conditions in mind a template is presented of the characteristics a teamwork evaluation should have.

Tab. 1. Summary of Teamwork Characteristics

Reasons for Working in Teams	**Teamwork Characteristic**	**Classroom Limitation**
System of system	Synergy among members	Semester-limited group history
Reduce intellectual workload-interdisciplinary teams	Group rewards	Individual grading
Reduce absolute workload	Formal organization and role definition	
Peer review-error checking	History of activity together	

These reasons/characteristics/limitations provide an extensive framework for developing a system of team for evaluating teamwork for accreditation. But it is probably impossible to identify a methodology that would adequately incorporate all of them. Instead methods the authors argue that any methodology used should try to couple characteristics from each of the first two columns with a constraint given in the second column. As long as the how the constraints affect the characteristics being measured are identified and are part of the evaluation, then it should be adequate. For example, having students from several departments participate in a multi-discipline design could be an effective means of evaluating teamwork. In such a situation, the design would be interdisciplinary, subject to peer review, provide synergy among members and, depending on how it was graded, some level of group rewards.

VII. CONCLUSIONS

Given that teams and teamwork is a requirement in today's multi-national, global business environment and classroom teamwork training must provide the principles and practices that provide a solid foundation for the student's future growth and training in this area. Groups which evolve out of the engineering classroom do not have the opportunity to work on teamwork principles and practices because of the requirements of the course. They are not exposed to other students from different disciplines in a true team setting and have limited time to understand the implications of teamwork. So, the selection of class projects should emphasize that a team (a collection of individuals who are *inter*dependent in their tasks, who *share responsibility* for outcomes, who see themselves and who are seen by others as an intact social entity) have to be measured collectively and individually. Introducing team skills early in the curriculum and that true laboratory teams be employed to simulate the team environment. The teams should be assessed on their success as a team as well as for course specific content and that the teaming assessment be included as part of the individual's grade.

As such, the authors also suggest building a project/curriculum template that measures the growth of an interdisciplinary team approach. This requires various disciplines setting up one or more interdisciplinary courses that succeed in achieving a true team environment. The selection of each course project emphases a team (a collection of individuals who are *inter*dependent in their tasks, who *share responsibility* for outcomes, who see themselves and who are seen by others as an intact social entity) that has grown to understand team assessment mechanisms. The characteristics for such course curriculum team considerations include several areas: [1]

The introduction and continual evolution into team understanding and evaluation throughout the student's development provides a better foundation and method for evaluating teamwork in an ABET review. Using a capstone teaming course that brings many of the key concepts together brings with it a true understanding of teaming and team performance measurement. This approach provides a better foundation and method for evaluating teamwork in an ABET review.

REFERENCES

[1] Perusich, K., Davis, B., Laware, G., and Taylor, K.: *Assessing Teamwork for Accreditation: Understanding what needs to be known and its integration into engineering and technology curricula*. 37th ASEE/IEEE Frontiers in Education, October 10-13, 2007, Milwaukee, WI

[2] Larson, C. & LaFasto, F., Teamwork: What must go right/What can go wrong. Beverly Hills, CA: Sage, Inc., 1999.

[3] LaFasto, F. & Larson, C., *When teams work best*. Thousand Oaks, CA: Sage, Inc., 2001.

[4] Kilmann, R., *Managing beyond the fix*, San Francisco, CA: Jossy-Bass Publishers, 1989.

[5] Benson, W. & Kappus, E., Managing People. *Your Competitive Edge in the 90's*, Emeryville, Ca: Avantos Performance Systems, Inc, 1992.

[6] Schuster, J.P., Carpenter, J., & Kane, M.P., *The Power of Open-Book Management*, New York: John Wiley & Sons, 1996.

[7] Tubbs, S.L..: *A Systems Approach to Small Group Interaction*. New York: McGraw-Hill Publishers. 2007.

Modelling Single Cell Electroporation with Bipolar Pulse: Simulating Dependance of Electroporated Fractional Pore Area on the Bipolar Field Frequency

Sadhana Talele
Department of Engineering
University of Waikato
Hamilton, New Zealand
Email: sadhana@waikato.ac.nz

Paul Gaynor
Department of Electrical and Computer Engineering
University of Canterbury
Christchurch, New Zealand

Jethro Van Ekeran
and Michael J. Cree
Department of Engineering
University of Waikato
Hamilton, New Zealand

***Abstract*—Electroporation EP, in which external electric field pulses create transient pores in a cell membrane, is an important technique for delivery of DNA and drugs into the cell. To enable entry of DNA into cells, the pores should have sufficiently large radii, remain open long enough for the DNA chain to enter the cell, and should not cause membrane rupture. A numerical model for a single spherical cell electroporated by application of direct and/or alternating external electric field pulses has been developed. The model is used to calculate the transmembrane potential, the number of pores and the the fraction of area occupied by the pores (fractional pore area *FPA*) in response to the various electric field pulses. Presented here are simulation results used to compare the extent of electroporation (fractional pore area *FPA*) in response to eletric field pulses of different frequencies in a range of extracellular conductivity for two cell raii. It is also observed that a 1 MHz bipolar sinusoidal applied electric field pulse reduces the relative difference in fractional pore area between the two cell sizes compared to a 100 kHz pulse.**

***Index Terms*—electroporation, AC pulses, pore density, numerical model, fractional pore area**

I. Introduction

Electroporation (EP - often also referred to as electropermeabilization) is now a common tool in biotechnology and is used in a number of medical treatments [1]–[3]. Until recently, the development of theoretical models of electroporation has lagged behind the experimental research. In order to optimize the efficiency of electroporation, it is important to consider as many biological and physical aspects as possible and it is a necessity that a variety of electric field pulse parameters be tried. Thus, a comprehensive model which can predict electropermeabilization as a result of any form of applied electric field pulse and other important electroporation parameters is necessary.

Previous reported studies [4], [5] model electroporation to calculate transmembrane potential V_m and the pore density (N). These models consider the non-linear behaviour of electroporation, which is due to the dynamics of membrane pore formation and its effect on the transmembrane potential in turn. However, these models assume a single non-varying pore radius around the cell membrane. Other models include spatial and temporal aspects of pore radius evolution [6], [7]. The electric fields used in these models were limited to unipolar DC pulses. The study presented here develops a model of a single spherical cell electroporation that simulates spatial and temporal aspects of pore radius as an effect of any given form of applied electric field (including unipolar or bipolar), and other important electroporation system parameters.

We consider two cell radii (15 μm and 7.5 μm) and calculate fractional pore area (*FPA*) to compare the extent of elctropermabilization at varying extracellular conductivities and the effect of higher frequency sinusoidal AC pulses.

II. Model of a single cell

Consider a spherical cell of radius a with intracellular conductivity σ_{in}. The cell is immersed in a medium with conductivity σ_{ex}. This system is exposed to a time-varying electric field $E(t)$. Azimuthal symmetry about the axis of applied electric field is assumed.

A. Transmembrane potential

It is assumed that the intracellular and extracellular regions are charge free. The homogeneous external electric field of strength $E(t)$ is used as a boundary condition, the potential being fixed at

$$\Phi(3a, \theta) = -3aE(t)\cos\theta, \tag{1}$$

on a sphere of radius $3a$ surrounding the cell, where θ is the polar angle. Since there are no sources inside the region, the potential obeys Laplace's equation

$$\nabla^2\Phi = 0, \tag{2}$$

except at the cell membrane at a radial distance of ($r = a$) where the potential is discontinuous because of the abrupt change in conductivity and the approximated infinitely thin membrane. Thus internal and external potentials can be defined as,

$$\Phi(r,\theta) = \begin{cases} \Phi_{\text{in}}(r,\theta) & r < a, \\ \Phi_{\text{ex}}(r,\theta) & a < r < 3a. \end{cases} \tag{3}$$

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_65,

The current across the membrane is used to relate the internal and external potentials [4], that is,

$$-\hat{\mathbf{r}}\cdot(\sigma_{\text{in}}\nabla\Phi_{\text{in}}(a,\theta)) = -\hat{\mathbf{r}}\cdot(\sigma_{\text{ex}}\nabla\Phi_{\text{ex}}(a,\theta)) = C_m\frac{\partial V_m}{\partial t} + J_m \quad (4)$$

where

$$V_m(\theta) = \Phi_{\text{in}}(a,\theta) - \Phi_{\text{ex}}(a,\theta). \quad (5)$$

Here $\hat{\mathbf{r}}$ is the unit outward radial vector, C_m is the specific membrane capacitance, V_m is the transmembrane potential and J_m is the current density at the cell membrane due to existing pores. The current density is made up of three terms as given in Equation 6

$$J_m = J_{\text{ion}} + J_{\text{sml}} + J_{\text{lge}}, \quad (6)$$

where $J_{\text{ion}} = g_m(V_m - V_{\text{rest}})$ is the ionic current density [4] (g_m is the specific membrane conductance, and V_{rest} is the membrane rest potential). The remaining two terms are explained in the following paragraphs.

J_{sml} is the current density through small pores and is given by

$$J_{\text{sml}} = N i_{\text{sml}}(r), \quad (7)$$

where N is the pore density of the initial small pores formed and $i_{\text{sml}}(r)$ is the diffusion current through a single pore of radius r (true for small pores only). A previously derived expression for i_{sml} [4], based upon the Nernst-Planck equation models, is used for pore radius below 1 nm, and is,

$$i_{\text{sml}} = \frac{\pi r^2\sigma_{\text{ps}}\nu_m RT}{Fh}\cdot\frac{(e^{\nu_m}-1)}{(G_-e^{\nu_m}-G_+)} \quad (8)$$

with

$$G_\pm = \frac{w_0 e^{w_0\pm n\nu_m}\pm n\nu_m}{w_0\pm n\nu_m}. \quad (9)$$

Here σ_{ps} is the conductivity of the aqueous solution that fills the pore (approximated by $\sqrt{\sigma_{\text{in}}\ \sigma_{\text{ex}}}$), F is Faraday's constant, R is the universal gas constant, T is the absolute temperature, h is the thickness of the membrane, w_0 is the energy barrier inside the pore, n is the relative entrance length of the pore and v_m is the nondimensional transmembrane potential [4] given by

$$v_m = V_m\left(\frac{F}{RT}\right). \quad (10)$$

This equation for pore current i_{sml} (equation 8) accounts for the electrical interactions between the ions and the pore wall [4]. Now, assume Q larger pores exist, and i_{lge} is the current through the electropores of radius larger than 1 nm. Then, J_{lge} is the total current density through Q larger pores; r_q being the radius of the q^{th} pore. Hence,

$$J_{\text{lge}} = \frac{1}{A}\sum_{q=1}^{Q} i_{\text{lge}}(r_q) \quad (11)$$

where A is the corresponding cell surface area. For these larger pores, the current–voltage relationship assumes that the transmembrane potential V_m occurs across the sum of pore resistance R_p and the series input resistance R_{in} [6], [8], as follows:

$$i_{\text{lge}}(r) = \frac{V_m}{R_p + R_{\text{in}}}, \quad (12)$$

where

$$R_p = \frac{h}{\pi\sigma_{\text{ps}}r_q^2}, \quad (13)$$

and

$$R_{\text{in}} = \frac{1}{2\sigma_{\text{ps}}r_q}. \quad (14)$$

B. Formation of pores:

Initially pores are assumed to be formed with the minimum-energy radius r_m = 0.76 nm, at a rate given by [4]

$$\frac{dN}{dt} = \psi e^{(V_m/V_{\text{ep}})^2}\left(1 - \frac{N}{N_{\text{eq}}(V_m)}\right), \quad (15)$$

where N is the pore density of the initial small pores formed, ψ is the creation rate coefficient and N_{eq} is the equilibrium pore density for a voltage V_m given by

$$N_{\text{eq}}(V_m) = N_0 e^{b(V_m/V_{\text{ep}})^2}. \quad (16)$$

Here, N_0 is the initial pore density with no applied electric field, V_{ep} is the characteristic voltage of electroporation and b is the pore creation constant equal to $(r_m/r_*)^2$, where r_* is the minimum radius of hydrophilic pores and r_m the minimum-energy radius [9]. All parameter values are as given in Table 1.

C. Evolution of pore radii:

The pores that are initially created with minimum-energy radius r_m change in size. Evolution of the radius of a pore is governed by the following equations 17 - 21 [6]. The lipid bilayer energy w_m depends on a number of parameters and is given by

$$w_m = \sum_{q=1}^{Q}\left[w_{\text{st}}\left(\frac{r_*}{r_q}\right)^4 + 2\pi w_{\text{ed}}r_q - \pi\xi_{\text{eff}}(A_p)r_q^2 + \int_0^{r_q}F(r_q,V_m)dr\right]. \quad (17)$$

The terms in this equation are explained in the following paragraphs.

In the first term, w_{st} is the steric repulsion energy [9]. The lipid head groups which line the pore interior, tend to repel each other due to steric and/or electrostatic interactions [10]–[12] and are taken into account by this term in equation 17.

The appearance of a circular pore in a membrane is balanced by the presence of two energy terms: reduction in energy barrier proportional to removal of pore area $\pi {r_q}^2$ (third term in equation 17) and increase in energy barrier by a linear edge component proportional to pore edge of length $2\pi r_q$ [13]–[15] (second term in equation 17). Here, w_{ed} is the pore edge energy, and ξ_{eff} the effective tension of the membrane is given by

$$\xi_{\text{eff}}(A_p) = 2\xi' - \frac{2\xi' - \xi_0}{\left(1 - \frac{A_p}{A}\right)^2} \quad (18)$$

where ξ' is the energy per area of the hydrocarbon-water interface [9], [11], ξ_0 is the tension of a membrane without pores and A is the total area of the lipid bilayer. A varying value of the total area A_p occupied by the pores at any given time is given by

$$A_p = \sum_{q=1}^{Q} \pi r_q^2 \tag{19}$$

and contributes to a dynamic effect of pore formation both temporally and spatially.

The last term in Equation 17 is the contribution of the membrane potential to the bilayer energy [9]. Assuming the inner surface of a pore as toroidal [9], [16], the electric force F acting on the pore is given by

$$F(r, V_m) = \frac{F_{max}}{\left(1 + \frac{r_h}{r + r_t}\right)} V_m^2. \tag{20}$$

This equation is a heuristic approximation [9] of the numerical solution which the authors have computed for the electrical force acting on a pore derived from first principles; Here, r_h and r_t are constants taken from reference [9]. This equation is thought to be appropriate for larger pores as it predicts that F approaches a constant value F_{max} as the pore radius increases, rather than increase linearly [15], or decrease to zero [17]–[19] as radius increases. All parameter values are as given in Table 1.

The rate of change of pore radii is given by [6]

$$\frac{dr_q}{dt} = -\frac{D}{kT} \frac{\partial w_m}{\partial r_q}, \qquad q = 1, 2, ..., Q \tag{21}$$

where D is the diffusion coefficient of the pore radius, k is Boltzman's constant and T is the absolute temperature.

III. Numerical Implementation

The model described above is implemented in MATLAB. The following description of model geometry is illustrated in Figure 1. Azimuthal symmetry about the axis of the applied electric field is assumed. the membrane is divided into a number of individually modelled slices (25 including the two poles). Each section of cell membrane is assigned a pore density N, of 'small pores', idealised as having equal radius r_m. Each section also has a transmembrane voltage (V_m) calculated. All results are simulated for a radial discretization step of $3a/51$.

Depending on the parameter values, a high number of pores may be formed. The total simulation time depends on the number of large pores created by the applied electric pulse. Strong pulses may create more pores, thus requiring long simulation times. To speed up computation in the event of such large numbers of pores forming, pores may be 'created in groups'. A single radius variable describes the radius of each pore in a group, thus reducing the number of radius variables required.

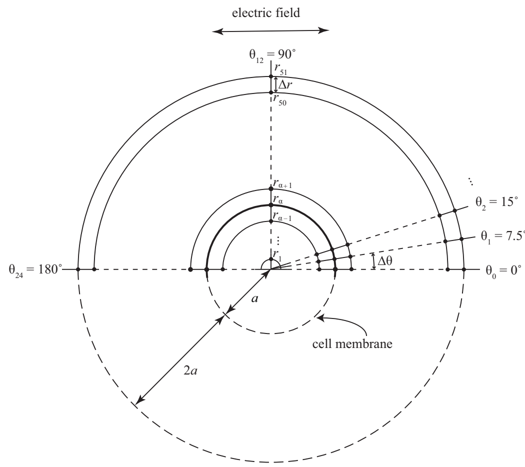


Fig. 1. Spherical cell model geometry showing the angle and radius discretization used in the numerical model

IV. Simulation and Results

Owing to the variable nature of biological growth dynamics, all electroporation applications treat cells that have some range of cell size. Although this range may only be in the order of a few percent, some applications may have ranges exceeding 200 percent [20]. As such, whatever the application, there may be a substantial benefit of being able to normalize the degree of electropermeabilization (considered equivalent to fractional pore area (*FPA*)) with respect to cell radius/size. Earlier passive and dynamic modelling of bipolar electric field induced transmembrane potential (V_m), has indicated that as the frequency of a constant amplitude applied electric field is increased, transmembrane potential (V_m) reduces and becomes less dependent on cell radius [5], [21]. In order to achieve a transmembrane potential (V_m) high enough for electroporation, the electric field amplitude must be increased at higher frequencies. To determine whether this effect translates to dynamic electroporation modelling that includes pore radii variability, two higher frequency electric field pulses of 100 kHz and 1 MHz were used for electroporation simulation. Two substantially (yet realistically) different cell radii of 15 μm, and 7.5 μm were considered. Simulations are carried out for evolution of transmembrane potential and pore radius for two applied bipolar electric field pulses at two significantly different frequencies to single spherical cell of two different radii. Pore radii and number of pores formed around a cell eventually decide the amount of drug intake in an electroporated cell. Both of these also vary with time. Thus, to compare the extent of electroporation, fractional pore area (*FPA*) is used here and results presented.

Figures 2 and 3 show fractional pore area (*FPA*) for 7.5 μm and 15 μm radius cells exposed to a two-cycle sine wave

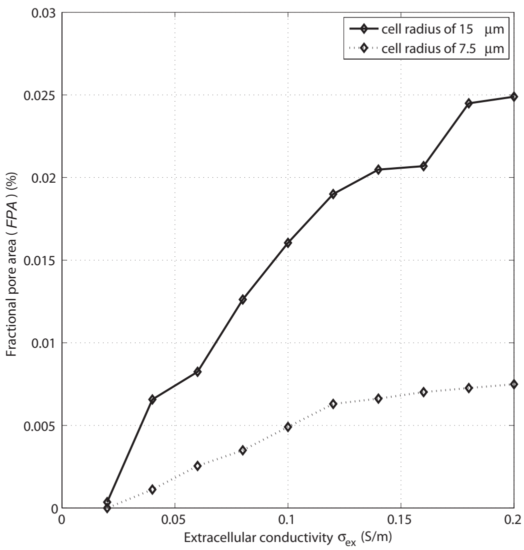


Fig. 2. Fractional pore area versus extracellular conductivity for 7.5 μm, and 15 μm cell radius exposed to a two-cycle sinusoidal bipolar applied electric field pulse of 130 kV/m peak magnitude at 100 kHz

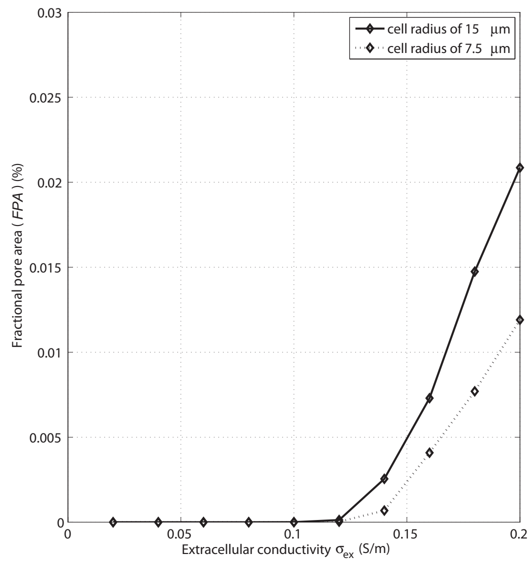


Fig. 3. Fractional pore area versus extracellular conductivity for 7.5 μm, and 15 μm cell radius exposed to a two-cycle sinusoidal bipolar applied electric field pulse of 350 kV/m peak magnitude at 1 MHz

electric field pulse of 100 kHz and 1 MHz respectively. The peak amplitude of the electric field, E_p was set to provide an average terminal fractional pore area (*FPA*) (for the two cell radii) of approximately 0.015% (presented in [22] as being a good level of permeabilization) at an extracellular conductivity (σ_{ex}) of 0.2 S/m. It is seen from Figures 2 and 3 that the relative difference in fractional pore area (*FPA*)) between the two cell sizes at extracellular conductivity (σ_{ex}) of 0.2 S/m is 1.1 for 100 kHz and 0.6 for 1 MHz. It is also observed that a 1 MHz bipolar sinusoidal applied electric field pulse reduces the relative difference in fractional pore area between the two cell sizes compared to a 100 kHz pulse.

V. Discussion

It is evident from Figures 2 and 3 that, considering 0.015 % fractional pore area (*FPA*) as the desired fractional pore area, two-cycles of 1 MHz sinusoidal bipolar electric field reduces the relative difference in fractional pore area (*FPA*) between the two cell sizes compared to two-cycles of a 100 kHz bipolar electric field. However, a significantly higher amplitude is required to create the same level of average fractional pore area (*FPA*). Lower values of extracellular conductivity (σ_{ex}) are warranted for better normalization of the degree of electroporation (*FPA*), as higher values of extracellular conductivity (σ_{ex}) increase the relative difference in fractional pore area (*FPA*) for cells of different radii/size.

VI. Conclusion

The fractional pore area tend to be more normalized when a 1 MHz bipolar sinusoidal field pulse is used as compared to a 100 kHz bipolar sinusoidal electric field for the same number of cycles if the external medium conductivity is not too high.

a	15.0 (μm)	cell radius
C_m	10^{-2} (Fm^{-2})	specific membrane capacitance [7]
h	5.0 (nm)	membrane thickness [4], [23], [24]
g_m	1.9 (Sm^{-2})	specific membrane conductance [4]
V_{rest}	–80 (mV)	membrane rest potential [4]
σ_{in}	0.3 (S m^{-1})	intracellular conductivity [23]
σ_{ex}	1.2 (Sm^{-1})	extracellular conductivity [24]
r_*	0.51 (nm)	minimum radius of hydrophilic pores [6]
r_m	0.8 (nm)	minimum energy radius at V_m=0 [6]
T	295 (K)	absolute room temperature [4]
n	0.15	relative entrance length of pores [4]
b	2.46	pore creation constant [4]
V_{ep}	258 (mV)	characteristic voltage of electroporation [4]
N_0	1.5×10^9 (m^{-2})	initial pore density [4]
w_0	2.65	energy barrier within pore [4]
ψ	1×10^9 (m^{-2} s^{-1})	creation rate coefficient [4]
w_{st}	1.4×10^{-19} (J)	steric repulsion energy [6]
w_{ed}	1.8×10^{-11} (J m^{-1})	edge energy [6]
ξ_0	1×10^{-6} (J m^{-2})	tension of the bilayer without pores [7]
ξ'	2×10^{-2} (J m^{-2})	tension of hydrocarbon-water interface [4]
F_{max}	0.70×10^{-9} (N V^{-2})	max electric force for V_m = 1 V [4]
r_h	0.97×10^{-9} (m)	constant [6]
r_t	0.31×10^{-9} (m)	constant [6]
D	5×10^{-14} (m^2 s^{-1})	diffusion coefficient for pore radius [7]

TABLE I
Geometric, electric and electroporation parameters used in simulation.

REFERENCES

[1] M. Cemazar, D. Miklavcic, and G. Sersa, "Intrinsic sensitivity of tumor cells to bleomycin as an indicator of tumor response to electrochemotherapy," *Japanese journal of cancer research*, vol. 89, pp. 328–333, 1998.

[2] G. Sersa, "Electrochemotherapy," in *Electrochemotherapy, Electrogenetherapy, and Transdermal Drug Delivery : Electrically Mediated Delivery of Molecules to Cells (Methods in Molecular medicine)*, M. J. Jaroszeski, R. Heller, and R. Gilbert, Eds. Totowa, New Jersey: Humana Press, 2000, ch. 6, pp. 119–133.

[3] G. Sersa, S. Novakovic, and D. Miklavcic, "Potentiation of bleomycin antitumor effectiveness by electrotherapy," *Cancer letters*, vol. 69, pp. 81–84, 1993.

[4] K. A. DeBruin and W. Krassowska, "Modeling electroporation in a single cell. I. Effects of field strength and rest potential." *Biophysical Journal*, vol. 77, pp. 1213–1224, 1999.

[5] S. Talele and P. Gaynor, "Nonlinear time domain model of electropermeabilization: Response of a single cell to an arbitrary applied electric field," *Journal of Electrostatics*, vol. 65, pp. 775–784, 2007.

[6] K. C. Smith, J. C. Neu, and W. Krassowska, "Model of creation and evolution of stable electropores for dna delivery," *Biophysical Journal*, vol. 86, p. 28132826, 2004.

[7] W. Krassowska and P. D. Filev, "Modeling electroporation in a single cell," *Biophysical Journal*, vol. 92, pp. 404–417, 2007.

[8] J. Newman, "Resistance for flow of current to a disk," *Journal of The Electrochemical Society*, vol. 113, pp. 501–502, 1966.

[9] J. C. Neu and W. Krassowska, "Modeling postshock evolution of large electropores," *Physical review E*, vol. 67, pp. 1–12, 2003.

[10] J. C. Weaver and Y. Chizmadzhev, "Theory of electroporation: a review," *Bioelectrochemistry and bioenergetics*, vol. 41, pp. 135–160, 1996.

[11] J. Israelachvili, *Intermolecular and Surface Forces*. London, UK: Academic Press, 1992.

[12] J. C. Neu and W. Krassowska, "Asymptotic model of electroporation," *Physical review E*, vol. 59, pp. 3471–3482, 1999.

[13] J. C. Weaver and K. T. Powell, "Theory of electroporation," in *Electroporation and Electrofusion in Cell Biology*, E. Neumann, A. E. Sowers, and C. A. Jordan, Eds. New York: Plenum Press, 1989, ch. 7, pp. 111–126.

[14] R. W. Glaser, S. L. Leikin, L. V. Chernomordik, V. F. Pastushenko, , and A. I. Sokirko, "Reversible electrical breakdown of lipid bilayers: formation and evolution of pores," *Biochim Biophys Acta*, vol. 940, p. 275287, 1988.

[15] I. G. Abidor, V. B. Arakelyan, L. V. Chernomordik, Y. A. Chizmadzhev, V. F. Pastushenko, and M. R. Tarasevich, "Electric breakdown of bilayer membranes: 1. the main experimental facts and their qualitative discussion," *Bioelectrochemistry and bioenergetics*, vol. 6, pp. 37–52, 1979.

[16] M. Kanduser, M. Fosnaria, M. entjurcb, V. Kralj-Iglic, H. Hgerstrandd, A. Iglia, and D. Miklavich, "Effect of surfactant polyoxyethylene glycol (c12e8) on electroporation of cell line dc3f," *Colloids and Surfaces A*, vol. 214, p. 205217, 2003.

[17] A. Barnett and J. C. Weaver, "Electroporation: a unified, quantitative theory of reversible electrical breakdown and mechanical rupture in artificial planar bilayer membranes," *Bioelectrochemistry and bioenergetics*, vol. 25, pp. 163–182, 1991.

[18] R. P. Joshi, Q. Hu, K. H. Schoenbach, and H. P. Hjalmarson, "Improved energy model for membrane electroporation in biological cells subjected to electrical pulses," *Physical review E*, vol. 65, no. 4, pp. 041 920–041 928, Apr 2002.

[19] V. F. Pastushenko and Y. A. Chizmadzhev, "Stabilization of conducting pores in blm by electric current," *Gen. Physiol. Biophys*, vol. 1, pp. 43–52, 1982.

[20] K. R. Foster and H. P. Schwan, "Dielectric properties of tissues," in *Handbook of biological effects of electromagnetic fields*, C. Polk and E. Postow, Eds. New York: CRC Press, 1996, ch. 1, pp. 25–97.

[21] C. Grosse and H. P. Schwan, "Cellular membrane potentials induced by alternating fields," *Biophysical Journal*, vol. 66, pp. 1632–1642, 1992.

[22] M. Hibino, M. Shigemori, H. Itoh, K. Nagayama, and K. J. Kinosita, "Membrane conductance of an electroporated cell analyzed by submicrosecond imaging of transmembrane potential," *Biophysical Journal*, vol. 59, pp. 209–220, 1991.

[23] T. Kotnik, D. Miklavcic, and T. Slivnik, "Time course of transmembrane voltage induced by time-varying electric fields a method for theoretical analysis and its application," *Bioelectrochemistry and bioenergetics*, vol. 45, pp. 3–16, 1998.

[24] T. Kotnik and D. Miklavcic, "Theoretical evaluation of the distributed power dissipation in biological cells exposed to electric fields," *Bioelectromagnetics*, vol. 21, pp. 385–394, 2000.

A Web-based System to Support Group Idea Generation for Creative Problem Solving

Ken'ichi Minamino[a,b] and Tetsuo Kinoshita[b]
[a]Faculty of Software and Information Science, Iwate Prefectural University,
152-52, Sugo, Takizawa-mura, 020-0193 Japan
[b]Graduate School of Information Sciences, Tohoku University,
2-1-1, Katahira, Aoba-ku, Sendai-shi, 980-8577 Japan
minamino@iwate-pu.ac.jp, kino@riec.tohoku.ac.jp

***Abstract*— As described in this paper, we propose a Web-based system to support group idea generation. Our prior study created a method of group idea generation to develop each group member's idea through positive group communication. We introduce positive idea evaluation into the method to support its function. The system provides functions of a computerized paper and group communication that consists of idea evaluation and online chat. Using the computerized paper function, all members can input their ideas simultaneously: they can see the same paper. Furthermore, using the function of group communication, a member who evaluates an idea (i.e. evaluator) competes with others in finding good points of the evaluated member idea. Moreover, during a chat session, the evaluators then tell the evaluated member ways to improve it. While all members of a group concentrate on the screen, they can do group work for group idea generation. Results show that the proposed system is effective to support group idea generation: the group members were able to generate many ideas.**

***Index Terms*— Creative problem solving, Group idea generation, Idea evaluation, Online chat, Web-based system**

I. INTRODUCTION

Recently, with respect to university education, students' acquisition of the ability to apply their expertise to problem solving and the ability to produce designs satisfying social requirements is assigned great importance. Both involve creativity, as shown for accrediting Japanese Engineering Education programs using the Japan Accreditation Board for Engineering Education (JABEE) criteria [1]. Internationally, the Information Systems 2002 model curriculum (IS2002) includes the topic of creativity in several of its Body of Knowledge elements [2]. Furthermore, suggestions have been made as to how creativity might fit into an Information Technology (IT) curriculum [3].

As a practical method of such education, we created a method of group idea generation in our prior study to develop each group member's idea, having positive group communication [4]. In this study, we realize a web-based system to support its work. The proposed system enables them to generate ideas more effectively than using a sheet of paper. We introduce positive idea evaluation into the method. Furthermore, the system provides functions of idea evaluation and online chat, in addition to computerization of a sheet of paper [5].

The process of creative thinking includes both divergent thinking, which is the ability to generate multiple solutions, and convergent thinking, which is the ability to find a single solution [9],[10]. Methods of idea generation are classifiable into divergent methods, convergent methods, integrated methods, etc. Although many studies have investigated the use of computer support for group idea generation using these methods, most have specifically examined generation of excellent ideas as a group. For example, computer-aided brainstorming and brainwriting [6], [7] are application systems of divergent methods. However, our system specifically examines developing ideas of each group member. Our goal is that each member of a group learns how to generate a feasible idea, applying their expertise to problem solving through group interaction.

Connolly et al. [13] studied the effects of anonymity and an evaluative tone on computer-mediated groups using a group decision support system to perform an idea-generation task. Based on the results of that experiment, identified groups working with supportive confederates were most satisfied and had the highest levels of perceived effectiveness, but produced the fewest original solutions and overall comments. When a subject proposed an idea and received an approving comment, he or she tended to interpret this as a signal that the idea was adequate and complete, and that no more work on that idea, or subsequent idea, was required. To deal with such issues, we discuss a method of promoting group communication based on positive idea evaluation, whereby each member can understand other participants' ideas and find their good points; then they can tell the other members ways to make it better, or ways to improve bad or insufficient points, etc.

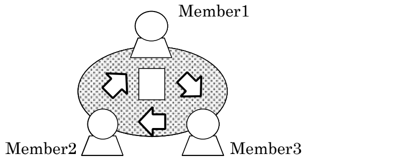
II. GROUP STUDY USING A METHOD OF GROUP IDEA GENERATION

A. Creative Problem Solving

For our prior study, we created a method of group idea generation. The method is useful in group study. This chapter presents a description of group study, tasks, and methods.

We chose tasks for each student such as defining specifications and developing software programs based on given conditions. Undergraduate students perform these tasks during comprehensive practice of Software Practice Class. The method is used in group study to generate ideas prior to the stage of defining specifications. A feasible idea is generated by

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_66,



	A	B	C
I	Initial idea that Member1 generated	Initial idea that Member2 generated	Initial idea that Member3 generated
II	Idea that Member2 developed from A-I	Idea that Member3 developed from B-I	Idea that Member1 developed from C-I
III	Idea that Member3 developed from A-II	Idea that Member1 developed from B-II	Idea that Member2 developed from C-II
IV	Final idea that Member1 generated	Final idea that Member2 generated	Final idea that Member3 generated

Fig. 1. Sheet of paper for the method in a prior study.

developing each member's initial idea in a group and having positive communication. Each member's initial idea is generated through the focus technique [10], [11]. After group study using the method, each member begins to define a specification based on the final idea; they develop software individually. An example of task is shown as the following.

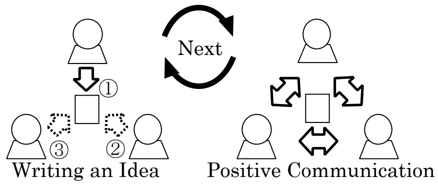
Create a character-based game that progresses by choosing an action from a list after every scene (e.g., role-playing game, character diagnosis).

Given conditions: (1) Create a game scenario comprising several scenes. When a player chooses an action, the player's state changes and the next scene is displayed. The game has multiple endings (e.g., good endings, bad endings); (2) Program a code using C language. Define a player state with C-structure and a scene with C-function. A scene function deals with input /output data, changes a structure variable of player's state, and calls the next scene function.

B. *Flow of Group Idea Generation*

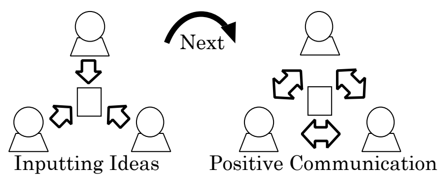
In the method used in our prior study, a group includes three members or more. All members of a group sit in proximity. A sheet of paper is used by the group. Each member develops other members' idea in turn through positive group communication: a member writes a developed idea for other members using positive communication; the member then passes the paper on to the next member in a group. Furthermore, the next member repeats it. The sheet of paper is presented in Fig. 1; the flow of the method is presented in Fig. 2. These figures are rules that apply to a group of three members. For a four-member group, the paper has five columns (I–V) and four rows (A–D). Although a four-member group can generate more ideas than a three-member group, it takes more time to finish the work.

Having positive communication for each idea, each member can understand other member's ideas and make it easy to develop them. We do not define a flow of positive communication and leave it to group's discretion. The result of the experiment in our prior study was that all students enjoyed the group study and generated final ideas. Then they defined specifications and were able to program using C language.



(1)Each member of a group writes an initial idea in the first row (I) of the column (A, B, C) of the sheet of paper in turn.
(2)Each member writes an idea in the second row (II) to develop the idea that was written in the upper cell by the other member, in turn.
At this time, all members have positive communication without criticism of each idea.
(3)Similarly each member writes an idea in the third row (III), giving positive communication, in turn.
(4)Each member writes a final idea that refers to the upper ideas from row I to row III.

Fig. 2. Flow of the method in a prior study.



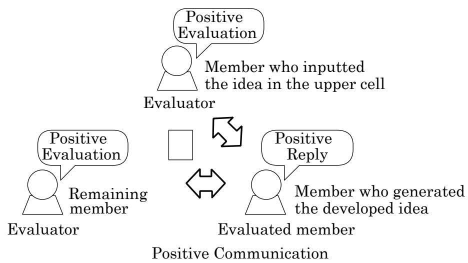
(1)Each member of a group inputs an initial idea in the first row (I) of the column (A, B, C) of the sheet of paper simultaneously.
(2)Each member inputs an idea in the second row (II) simultaneously, to develop the idea that was inputted in the upper cell by the previous member.
(3)All members start positive communication without criticism of each idea.
(4)Similarly each member inputs an idea in the third row (III) simultaneously, giving positive communication in turn.
(5)Each member inputs a final idea that refers to the upper ideas from row I to row III.

Fig. 3. Flow of the method using the computerized paper of the web.

III. WEB-BASED SYSTEM TO SUPPORT GROUP IDEA GENERATION

A. *Issue of Support for Group Idea Generation*

Because of computerization of the sheet of paper presented in Fig. 1 and allocation of one computer to every member, members of the group did not need to pass the paper on to the next member; they were able to concentrate their attention on it. In fact, as depicted in Fig. 3, it is possible to create a flow by which all members were enabled to input ideas simultaneously. However, as a result of experiments using only a computerized paper of the web, groups using the computerized paper exchanged fewer words than groups using the sheet of paper. They came to concentrate too much on the computerized paper. However, it is important to have positive communication so that each member can understand ideas of other members, identify their good points, and then tell the other members



(1)Evaluators evaluate the developed idea positively. They can ask questions if necessary.
(2)Evaluators tell the evaluated member about good points of the developed idea and ways to make it better (or ways to improve a bad or insufficient point, etc).
(3)In reply to the evaluations, the member tells evaluators about an impression of the evaluations, etc.

Fig. 4. Flow of Positive Communication.

ways to make it better, or ways to improve a bad or insufficient point, etc.

As described in this paper, we propose an approach to promote group communications that are helpful in generating and developing ideas. Therefore, we create the flow of positive communication based on idea evaluation and provide functions to support their performance.

B. Flow of Group Idea Generation using the Web

In addition to the computerization of a sheet of paper, we introduce positive idea evaluation into the method: we give members time to evaluate their idea to find their good points. Thereby, they share positive communication to tell each other ways to make it better, or ways to improve a bad or insufficient point, etc. The flow of positive communication is depicted in Fig. 3. In a group, each member has a role. An evaluated member and two evaluators form a three-member group. The evaluated member is the one who generated the developed idea. One evaluator is the person who inputted the original idea in the upper cell previously. Furthermore, the other evaluator is the remaining member: he or she is also a person charged with expediting the process. First, evaluators evaluate the developed idea positively; then all members participate in positive communication.

Because the flow of positive communication applies to each member's idea, they are repeated three times for three ideas that are written in step II or III of the sheet of paper shown in Fig. 1. In successful cases, members were able to generate many ideas like brainstorming sessions. Therefore, each member receives many ideas from other members in addition to the ideas inputted in the paper. These ideas would be helpful to them when they generate the developed ideas and the final idea.

C. Function of the Web-based System

The proposed system provides two functions: a function of a sheet of paper and a function of communication. Using the sheet of paper function, which provides computerized writing paper, all members can input their ideas simultaneously; they can see the same paper. For that reason, each member can input an idea while viewing the others' input ideas. When all members finish inputting their ideas, they click the Next button to proceed to the next step of the flow, as depicted in Fig. 3. The function is implemented in Perl, using CGI on the web server. In the computerized paper, only the part that each member can input as text is described by the html form.

The function of communication consists of idea evaluation and online chat. At the first step of the flow portrayed in Fig. 4, members that evaluate a developed idea (namely, evaluators) vote on how interesting it is. Points are scored by pushing an evaluation button. A member can push the button up to a maximum of 10 times per idea. All members can see the evaluators' points so that evaluators will push their evaluation button positively. Then, at the second and third step of the flow portrayed in Fig. 4, Members of a group communicate using the online chat. Although they have no strict time limit to the group communication for each developed idea, the tentative time limit is five minutes. If necessary, the group might use more than five minutes.

The function was implemented as a Java Servlet, which runs in a web server, and as a Java Applet, which runs in a web browser. The applet window will appear on the screen automatically when the members start positive communication. Similarly, when they return to the computerized paper of the web, the window will disappear from the screen. At that time, the chat logs (including the evaluation scores) will be displayed automatically in the computerized paper.

The network of the proposed system is presented in Fig. 5 and the structure of computerized paper is depicted in Fig. 6. Students use their ID and passwords to access the system.

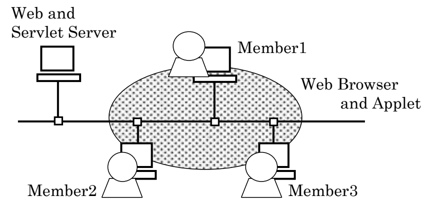


Fig. 5. Network of Web-based System.

	A	B	C
I	Initial idea that Member1 generated	Initial idea that Member2 generated	Initial idea that Member3 generated
II	Idea that Member2 developed from A-I	Idea that Member3 developed from B-I	Idea that Member1 developed from C-I
	Chat log of A-II	Chat log of B-II	Chat log of C-II
III	Idea that Member3 developed from A-II	Idea that Member1 developed from B-II	Idea that Member2 developed from C-II
	Chat log of A-III	Chat log of B-III	Chat log of C-III
IV	Final idea that Member1 generated	Final idea that Member2 generated	Final idea that Member3 generated

Fig. 6. Structure of Computerized Paper.

Fig. 7. Examples of the Screen of the Computerized paper and Applet window (Japanese version).

Students are assigned randomly and automatically to three-member groups or four-member groups. The administrator has only to set up the ID and password for the students using the administrator page. When they finish the work, they can save the computerized paper as an html file. Screen shots are portrayed in Fig. 7 (Japanese version). The left panel presents an example of the screen of the computerized paper of the web; the right panel shows the applet window. In the left panel, the computerized paper structure corresponds to the structure depicted in Fig. 6. In the right panel, the function of idea evaluation is located in the upper part of the window (i.e. an evaluation button and score indications). However, the evaluation button is disabled in the window of the evaluated member. Furthermore, the online chat function is located in the lower part of the window.

We expect that the function of idea evaluation using the applet window promotes group communication: the evaluator competes with others in finding good points of the evaluated member's idea. Furthermore, they tell the evaluated member ways to improve the idea during the chat session. Therefore, while all members of a group concentrate on the screen, they can do group work for group idea generation.

IV. Experiments

A. Methodology

We clarify the effectiveness of the functions of the proposed system. An experiment is conducted to compare the proposed systems with a system that includes no idea evaluation function (i.e. no evaluation button or score indications in the applet window). We investigate group communication and the quantity and quality of ideas that members of all groups generated. From the viewpoint of divergent thinking, the quantity and quality of ideas are evaluated based on fluency (Number of ideas), flexibility (Diversity of idea classifications) and originality (Uniqueness of the ideas) [8],[10]. Additionally, we investigate the games that students of subjects created after this experiment, as creative products. From the viewpoint of convergent thinking, we investigate the game specifications in the students' papers. The specifications are evaluated based on originality and sensibility, following Finke et al. [12].

In the experiment, 12 undergraduate students participated. The subjects were all first year students. They were assigned randomly to three-member groups. In all, four groups were formed. Half of all groups used the proposed system; the rest used another system. The subjects solved the task described in chapter II in the comprehensive practice of Software Practice Class at Faculty of Software and Information Science, Iwate Prefectural University: Create a character-based game that is progressed by choosing an action from a list after every scene.

We did not conduct an experiment to investigate the effect of group communication used by members who are temporally and spatially dispersed. Each member of the groups used a computer simultaneously and in the same room.

In this experiment assessing both systems, group members sat in proximity and used networked computers in our laboratory. All group members of both systems input a developed idea within five minutes to concentrate on it. Furthermore, they had a time limit of five minutes for group communication for each developed idea. However, in cases where some member was unable to generate an idea easily, they were allowed more time.

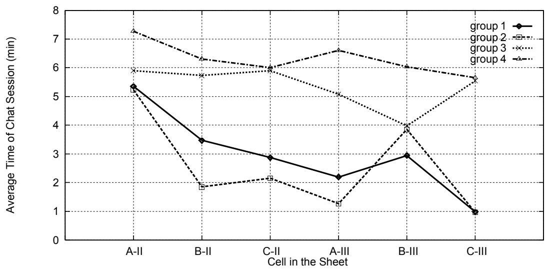


Fig. 8. Average time of chat session.

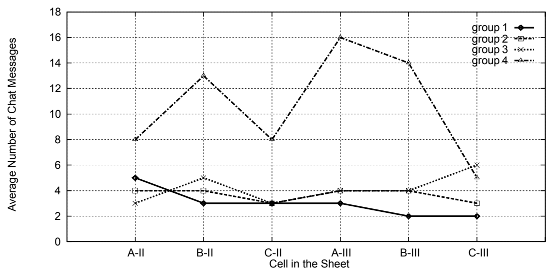


Fig. 9. Average number of chat messages.

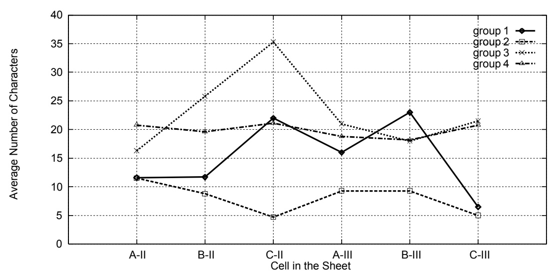


Fig. 10. Average number of characters in a chat message.

B. Comparison system

Another system, which was used for comparison, was the same as the proposed system except for its lack of the function of idea evaluation. The applet window had no evaluation button or score indications in the upper part of the applet window portrayed in Fig. 7. Within the group, the members were able to communicate using only the online chat function. The method flow shown in Figs. 3 and 4 were unchanged for both systems.

The proposed system enables evaluators to express their impressions using the idea evaluation function. The system used for comparison had no such function. Therefore, members had to express their impression using online chat. We investigated the effects of the idea evaluation function through group idea generation.

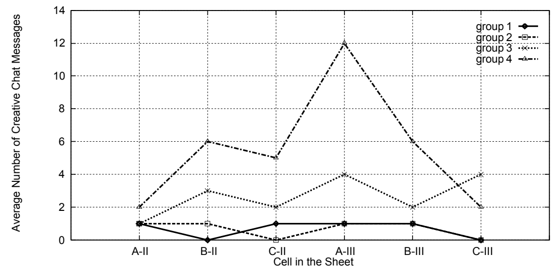


Fig. 11. Average number of creative chat messages.

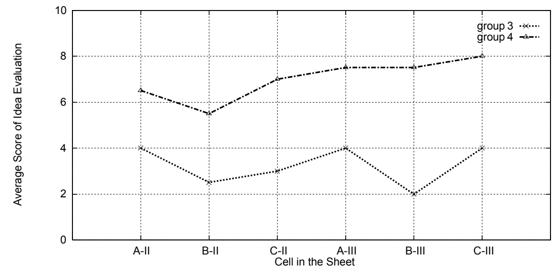


Fig. 12. Average score of idea evaluation.

V. RESULTS

A. Analysis and evaluation of group communication

Members of all groups were able to understand the flow and were able to generate their final ideas to solve the task. A number is allotted to each group to identify the groups. Groups 1 and 2 used the alternate system; groups 3 and 4 used the proposed system. Results of analyses of group communication are presented in Figs. 8, 9, 10, 11, and 12. These results are, respectively, the average times of chat sessions, the average numbers of chat messages, the average numbers of characters in chat messages, the average numbers of creative chat messages and the average scores of idea evaluations for cells in the computerized paper. The figures presented in Fig. 7 are screen shots of group 4's activities in this experiment.

As shown in the results presented in Fig. 8, the time in groups 3 and 4 is longer than the time in groups 1 and 2 (the average time in groups 1 and 2 is 2.8 min and the average time in groups 3 and 4 is 5.8 min). As shown in the result shown in Fig. 9, the number of messages in groups 3 and 4 is greater than the number of messages in groups 1 and 2 (the average number in groups 1 and 2 is 3.3 times; the average number in group 3,4 is 7.4 times). Results presented in Fig. 10 show that the characters used by groups 3 and 4 were more numerous than the characters in groups 1 and 2 (11.6 characters, average, in groups 1 and 2; 21.4 characters, average, in groups 3 and 4). As shown in Fig. 11, the creative messages in groups 3 and 4 were more numerous than the creative messages in groups 1 and 2 (0.7 times, average, in groups 1 and 2; 4.1 times, average, in groups 3 and 4). Furthermore, from the result presented in Fig.

12, members of groups 3 and 4 scored the developed ideas freely.

Members of groups 3 and 4 concentrated more on group idea generation than members of groups 1 and 2. The proposed system is more effective in supporting group communication than the system used for comparison. The proposed system provides the function of idea evaluation, which enables evaluators to express their impressions.

In addition, although group members proposed ideas and received approving comments, they did not interpret this as a signal that the idea was adequate and complete and that no more work on that idea was necessary. We were able to induce participants to find many good points related to the idea using an idea evaluation button; we also compelled them to tell each other ways to improve the idea during the chat session.

B. *Analysis and evaluation of ideas' quantity and quality*

From the viewpoint of the divergent thinking, we evaluated quantity and quality of ideas that members were able to generate using the two systems, based on fluency (number of ideas, except duplicate ideas, unfeasible ideas, and ideas irrelevant to the task), flexibility (number of classifications of ideas) and originality (number of unique ideas). In the evaluation of flexibility, ideas are classified into the following five classifications: (1) scenario composition, (2) data structure, (3) input (e.g. idea about the kind of action), (4) output (e.g. idea about the representation of scene), and (5) algorithm.

The results of quantity and quality of ideas are shown in Table I. The initial ideas and the final ideas in the paper are excluded from these analyses from the viewpoint of the divergent thinking. However, the creative chat message presented in Fig. 11 is included as an idea. From the result presented in Table I, in fluency, flexibility, and originality, the values of groups 3 and 4 are greater than the values of groups 1 and 2. The proposed system is more effective for supporting group communication than the system used for comparison.

C. *Evaluating Creative Products*

From the viewpoint of convergent thinking, the game specifications are rated on originality and sensibility using a five point scale (from 1 'very poor' to 5 'very good'). Both originality and sensibility were used to exclude specifications that were original but meaningless. Then, 15 undergraduate students, all second year students who had solved the same task the prior year, rated the specifications. They rated the specifications from their own experience.

The result of evaluating the creative products is presented in Table II, which shows specifications rated at greater than or equal to 3.0 on average, for both originality and sensibility. From the result described in Table II, the specifications in groups 3 and 4 are more numerous than the specifications in groups 1 and 2. Members of groups 3 and 4 created specifications based on final ideas that were generated using the proposed system. Groups 3 and 4 generated final ideas more effectively than groups 1 and 2.

TABLE I
QUANTITY AND QUALITY OF IDEAS

Group	Number of Ideas	Fluency	Flexibility	Originality
1	10	7	3	3
2	15	8	2	3
3	22	19	4	5
4	39	34	5	5

TABLE II
EVALUATING CREATIVE PRODUCTS

Group	Number of Scenes	Originality	Sensibility
1	20	3.5	3.0
2	9	3.0	3.1
3	9	3.5	3.7
3	14	3.5	4.1
4	7	3.5	4.0
4	43	4.1	4.7

VI. CONCLUSION

Using the proposed system, each member must do the following to develop each member's idea in a group: (1) Apply knowledge acquired during the Software Practice Class to develop other members' ideas; (2) Generate an idea from new viewpoints, while referring to other members' ideas and messages. The proposed approach demonstrably supported group idea generation: because it provides the idea evaluation function, the proposed system promotes idea generation and group communication.

Future research should be undertaken to study group idea generation, taking advantage of both the proposed system based on synchronous communication and a multi-agent system [14] based on asynchronous communication.

REFERENCES

[1] Japan Accreditation Board for Engineering Education (JABEE), "Guide to Criteria for Accrediting Japanese Engineering Education Programs," http://www.jabee.org/. (in Japanese)
[2] J. T. Gorgone, G. B. Davis, J. S. Valacich, H. Topi, D. L. Feinstein, and J. H. E. Longenecker, "IS 2002: Model curriculum and guidelines for undergraduate degree programs in information systems," http://www.is2002.org/.
[3] R.B. Sweeney, "Creativity in the information technology curriculum proposal," Proc. of the 4th conference on Information technology curriculum, pp.139-141, 2003.
[4] K. Minamino, K. Sekiguchi, A. Abe, and Y. Watanabe, "Group study using idea generation method for creative problem solving in the software practice," Journal of Information Processing Society of Japan, vol.47, no.5, pp.1578-1592, 2006.(in Japanese)
[5] K. Minamino, T. Terui, and T. Kinoshita, "Group Idea Generation Support System for Creative Problem Solving," IEICE Transactions on Information and Systems, vol.J91-D, no.2, pp.166-177, 2008. (in Japanese)
[6] K. Holt, "Brainstorming – from classics to electronics," Journal of Engineering Design, vol.7, no.1, pp.77-82, 1996.
[7] H. Kurose, "Cooperative Idea Generation System BWS-1," IPSJ SIG Technical Report 1993-GN-003, vol.1993, no.75, pp.57-64, 1993. (in Japanese)
[8] U. Sak and C.J. Maker, "Divergence and convergence of mental forces of children in open and closed mathematical problems," International Education Journal, vol.6, no.2, pp.252-260, 2005
[9] J.P. Guilford, The Nature of Human Intelligence, New York McGraw -Hill, 1967.
[10] M. Takahashi, The Bible of Creativity, JUSE Press, 2002. (in Japanese)

[11] C.S. Whiting, Creative Thinking, Reinhold Publishing Co., New York, 1958.
[12] R.A. Finke, T.B. Ward, and S.M. Smith, "Creative cognition: Theory, research, and applications," MIT press Cambridge, MA, 1992.
[13] T. Connolly, L.M. Jessup, and J.S. Valacich, "Effects of anonymity and evaluative tone on idea generation in computer-mediated groups," Management Science, vol.36, no.6, pp.689-703, 1990.
[14] K. Minamino and T. Kinoshita, "A Multi-Agent System to Support Group Idea Generation for Creative Problem Solving," Proc. of Joint 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on advanced Intelligent Systems (SCIS & ISIS 2008), 2008.

Unsupervised Segmentation of Industrial Images Using Markov Random Field Model

Mofakharul Islam John Yearwood Peter Vamplew
Center of Informatics and Applied Optimization
School of Information Technology and Mathematical Sciences
University of Ballarat,
Australia

Abstract- **We propose a novel approach to investigate and implement unsupervised image content understanding and segmentation of color industrial images like medical imaging, forensic imaging, security and surveillance imaging, biotechnical imaging, biometrics, mineral and mining imaging, material science imaging, and many more. In this particular work, our focus will be on medical images only. The aim is to develop a computer aided diagnosis (CAD) system based on a newly developed Multidimensional Spatially Variant Finite Mixture Model (MSVFMM) using Markov Random Fields (MRF) Model. Unsupervised means automatic discovery of classes or clusters in images rather than generating the class or cluster descriptions from training image sets. The aim of this work is to produce precise segmentation of color medical images on the basis of subtle color and texture variation. Finer segmentation of images has tremendous potential in medical imaging where subtle information related to color and texture is required to analyze the image accurately. In this particular work, we have used CIE-Luv and Daubechies wavelet transforms as color and texture descriptors respectively. Using the combined effect of a CIE-Luv color model and Daubechies transforms, we can segment color medical images precisely in a meaningful manner. The evaluation of the results is done through comparison of the segmentation quality with another similar alternative approach and it is found that the proposed approach is capable of producing more faithful segmentation.**

I. INTRODUCTION

Segmentation is the process that separates objects in an image. In medical images, the aim is to separate different parts of the anatomy. Segmentation plays a crucial role in automatic processing for analysis and evaluation of medical images. A reliable and faithful segmentation can bring enormous benefits in many medical imaging problems including image guided surgery, radiation therapy for cancer patients and identification of anatomical objects that are extremely difficult to visualize when no special contrast fluid is added to the blood. In addition, recognition of anatomical objects like joint surfaces, cysts and different brain parts would benefit enormously from reliable segmentation. Image segmentation involves separating objects from the background by partitioning the image into units that are identical in respect of one or more features. Image segmentation is obviously a crucial task in medical imaging. However, manual segmentation is not only a tedious job but time consuming as well. Further, most of the time it is inaccurate as manual segmentation by experts has been shown to be variable. So, there is a need for algorithms that are accurate and require human interaction as little as possible.

Using a finite mixture model in color image segmentation is a popular approach as in statistical model-based methods, computationally intensive algorithms are a common drawback. As a result, finite mixture models have received significant attention from researchers in the last decade. Finite mixture model has been further modified by Sanjay-Gopal and Herbert [1] to Spatially Variant Finite Mixture Model (SVFMM) by introducing a process based on Markov Random Fields (MRF) model, which is able to capture the spatial relationship among neighboring pixels. In doing so, it can efficiently handle specific situations like recognition of anatomical objects from the endoscopic, colonoscopic, and echo color dopler as these images are usually associated with numerous artifacts such as noise, partial volume effects, and intensity inhomogeneity.

From an image segmentation point of view, approaches based on the SVFMM as proposed in other works [1], [4] have a common limitation that it can handle single feature only. The reason is, the SVFMM is a single dimensional model. As a result, its application is limited to gray level images only. In our proposed approach, we have improved and modified the model to make it efficient for longer set of features, which makes it enable to handle multiple feature set. Therefore, our proposed Multidimensional Spatially Variant Finite Mixture Model (MSVFMM) is capable to handle multiple cues including color and texture features simultaneously.

The remainder of this paper proceeds as follows: In Section II, we describe the proposed Multidimensional Spatially Variant Finite Mixture Model (MSVFMM) that we will use in our proposed approach for image segmentation. In Section III, we present our novel approach followed by Section IV, where we present previous work. In Section V, we present some of the significant features of our proposed approach. Experimental results demonstrating the accuracy and effectiveness of the proposed approach are discussed in Section VI and finally in Section VII, we present our conclusion and future work.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_67, © Springer Science+Business Media B.V. 2010

II. MULTIDIMENSIONAL SPATIALLY VARIANT FINITE MIXTURE MODEL (MSVFMM)

From an image segmentation point of view, approaches based on the SVFMM as proposed in other works [1], [4] have a common limitation that it can handle single feature only. The reason is, the SVFMM is a single dimensional model. As a result, its application is limited to gray level images only. In our proposed approach, we have improved and modified the SVFMM into a full functional MSVFFM to make it efficient for longer set of features, which makes it enable to handle multiple feature set. Therefore, our proposed multidimensional SVFMM is capable to handle multiple cues including color and texture features simultaneously and efficiently. Further, we have integrated the number of components finding tool and features extraction tools into the model to make it as a robust color image segmentation approach that eventually can handle a wide variety of natural and/or industrial color images.

Basically, SVFM model is a classical finite mixture model that employs MRF model to capture the neighborhood relationships among the neighboring pixels. We have proposed a *multidimensional spatially variant finite mixture model* (MSVFMM) which is able to handle a large feature vector rather than a single feature as proposed by Sanjay-Gopal & Herbert and Blekas et al., which are in fact, single dimensional [1], [4]. In statistical model-based methods, computationally intensive algorithms are a common drawback. So, finite mixture model in image segmentation is a popular approach. As a result, classical finite mixture models have received significant attention from researchers in the last decade. Finite mixture model has been further modified by Sanjay-Gopal and Herbert [1] and Blekas et al [4] to *spatially variant finite mixture models* (SVFMM) by introducing a process based on Markov Random Fields (MRF), which is able to capture the spatial relationship among neighboring pixels. So, SVFMM has some advantageous features, especially in the case of contextually dependent physical phenomena like images as it incorporates the outstanding features of the MRF model without additional computational burden. In the proposed MSVFMM method, MRF introduces a prior distribution that takes into account the neighborhood dependency or relationship among the neighboring pixels. We elaborate a stochastic model based segmentation framework based on Gaussian Mixture Models, which constitute a well-known probabilistic Neural Network Model.

In fact, the SVFMM model is a modification of the classical mixture model or Gaussian mixture model with the introduction of MRF model. We assume a mixture model with an unknown number of components K each one having its own vector of density parameter θ_j for each feature.

In our multidimensional model, each ith pixel of the image is represented as a vector of values as $\chi^i = (x_1^i, x_2^i, ..., x_n^i)$ where x_n^i denotes pixel values in the nth dimensional plane. So, the probabilities of the ith pixel belonging to the jth class label are given as;

$$\eta_j^i = P(j \mid x^i) \tag{1}$$

where, x^i = observation at the ith pixel. Here, model parameters should strictly satisfy several constraints at the end of this section.

Similar to the SVFMM, we assume that the density function $f(x_i \mid \Phi, \Gamma)$ at any observation x_i is given by:

$$f(x_i \mid \Phi, \Gamma) = \sum_{j=1}^{K} p_j^i \psi(x_i \mid \theta^j) \tag{2}$$

where, $\psi(x_i \mid \theta^j)$ is a Gaussian distribution with parameters $\theta_j = (\mu_j, \sigma_j)$, and K is the number of components in the mixture.

Maximum a posteriori (MAP) introduces a prior distribution for the parameter set Φ that takes into account spatial information based on the *Gibbs function*. According to the Hammersley-Clifford theorem, *Gibbs distribution* takes the following form;

$$P(\Phi) = \tfrac{1}{Z} \exp(-U(\Phi)) \tag{3}$$

where, $U(\Phi) = \beta \sum V_{N_i}(\Phi)$.

Φ is a vector of features. The function $-U(\Phi)$ is an energy term. β is called the *regularization parameter*, the normalizing constant Z is called a partition function and $V_{N_i}(\Phi)$ denotes *clique potential* of the label configuration p^m within the neighborhood N_i of the ith pixel which can be calculated as;

$$V_{N_i}(\Phi) = \sum_{m \in N_i} g(u_{i,m}) \tag{4}$$

Where the $u_{i,m}$ denotes the distance between the two label vectors p^i and p^m. The function *g(u)* must be gradually increasing and nonnegative.

A posterior log density function can be derived from (3) as;

$$P(\Phi, \Gamma \mid X) = \sum_{i=}^{N} \log(x^i \mid \Phi, \Gamma) + \log P(\Phi) \tag{5}$$

The *Expectation-Maximization* (EM) algorithm requires that the computation of the conditional expectation value z_j^i of the hidden variables must be at the *Expectation* step due to MAP estimation of the parameters $\{p_j^i\}$ and $\{\theta_j\}$;

$$z_j^{i^{(t)}} = \frac{p_j^{i^{(t)}} \psi(x^i | \theta_j^{(t)})}{\sum_{l=1}^{K} p_l^{i^{(t)}} \psi(\chi^i | \theta_j^{(t)})} \tag{6}$$

Then, maximization of the following log-likelihood corresponding to the complete data set is performed in the *Maximization* step;

$$Q_{MAP}(\Phi, \Gamma \mid \Phi^{(t)}\Gamma^{(t)}) \quad (7)$$

where, t = iteration step.

For each parameter, Q_{MAP} can be maximized independently, which yields the update equations for parameter of the component densities $\mu_j^{(t+1)}$ and $[\sigma_j^2]^{(t+1)}$. However, maximization of the function Q_{MAP} with respect to the label parameter p_j^i does not provide a closed form of update equations. In addition, the maximization procedure must also take into the constraints $0 \le p_j^i \le 1$ and $\sum_{j=1}^{K} p_j^i = 1$. However, this difficulty has been encountered successfully by Sanjay-Gopal & Herbert [1] and subsequently by Blekas et al. [4]. Sanjay-Gopal and Herbert [1] has introduced a Gradient Projection (GP) algorithm to make the EM algorithm using closed form expressions while Blekas et al [4] implemented the M-step based on a closed form update equation followed by an efficient projection method.

III. Proposed Color Image Segmentation Approach

Conceptually, in most cases it is desired that the same cluster label be assigned to spatially adjacent pixels. In implementing these ideas, the Bayesian framework provides a natural approach. Pixel intensity information is captured by the likelihood term in Bayesian Networks, while a prior biasing term captures the spatial location information with the help of MRF. Due to the complexity of determining color distance in RGB spaces, we have used perceptually uniform CIE-Luv color space. Considering the advantages and disadvantages of different texture descriptors, we have chosen Daubechies transform, a Discrete Wavelet Transform (DWT), where wavelet analysis breaks up an original signal into shifted and scaled versions. Moments of wavelet coefficients in various frequency bands have been shown to be effective for representing subtle variation in texture [16]. We have implemented a completely unsupervised image segmentation technique for color textured images in our present work, which uses SVFMM and Markov Random Field Model. We can use the multivariate normal density for classification of pixels as it is typically an appropriate model for classification where the feature vectors are continuous valued mildly corrupted versions of a single mean vector [5]. Pixels from this type of distribution tend to cluster about the mean, and the variance will determine up to what extent they will spread out.

A. Color Feature Extraction

Choice of a suitable color space is a crucial aspect of color feature extraction. The most common color space in computer graphics is RGB, which is a three dimensional color space. Due to the lack of equal perception of color dissimilarity in RGB, there is a need to non-linear transform of the RGB color space in order to develop a perceptually uniform model of color space. Among the transformed color spaces, HSV, CIE-Lab and CIE-Luv are most common. We prefer CIE-Luv as it represents the three characteristics i.e., hue, lightness and saturation that characterize the perceptually uniform color efficiently.

B. Texture Feature Extraction

Texture characterizes local variations of pixel's color or intensity. There is no formal or unique definition of texture though texture based methods are commonly being used in computer vision and graphics. Each texture analysis method defines texture from its own perspective. Texture can be defined as a local statistical pattern of texture primitives in observer's domain of interest. Texture is often describe as consisting of primitives that are arranged according to placement rule which gives rise to structural methods of texture analysis that explicitly attempts to recover the primitives and the replacement rules. On the other hand, statistical methods consider texture as random phenomena with identifiable local statistics. The neighborhood property is a common similarity among all the texture descriptions. So, we can define texture as an image feature which is characterized by the gray value or color pattern in a neighborhood surrounding the pixel.

We can use many texture extraction tools or descriptors to extract texture feature. Using Wavelet transform is a recent trend in image processing. Although other texture descriptors like Gabor and MRSAR filters [10], [15] are popular but from the computational burden point of view, wavelet transform performs better than other texture descriptors. Gabor filters are found efficient in capturing the strong ordered or coarser texture information while MRSAR filters works well in capturing weak ordered or finer texture details. Wavelet transform, on the other hand, is found to be efficient in capturing both the coarser and finer texture information simultaneously with lesser computational burden.

Having been inspired by Islam et al's [8] work, where they used Haar wavelet transform and found promising results in capturing texture information, we have decided to utilize another member of wavelet family i.e., Daubechies wavelet transform, which is more accurate in discriminating texture. The Daubechies wavelet transform is named after its inventor, the mathematician Ingrid Daubechies. The Daubechies D4 transform has four wavelet and scaling function coefficients.

C. Finding The Number of Components

We have tried several data clustering approaches to identify the number of components automatically but most of them have been found inappropriate for handling image data properly. A few of them are found incapable to handle data having corrupted values as many pixels may not have true values for some attributes while few of them are not performing well in handling discrete data like image data.

SNOB [2], [3], a *Minimum Message Length* (MML) based unsupervised data clustering approach is free from this sort of difficulties and has been found more effective in handling noisy color image data that have a diverse varieties of faults and imperfections.

In general SNOB has been found more or less effective in identifying the number of components in an image but sometimes it produces spurious results due to its limitations to handle some imaging artifacts like noise, shadow, and intensity inhomogeneity. In order to remove this difficulty of SNOB, we utilize the technique *Ensembles of Clusterings*. We use SNOB as a starting point for getting multiple clusterings based on which Ensembles of Clusterings produce a consolidated clustering. The procedure we have followed for determining the number of components is as follows: produce multiple clusterings for different parameter values and select the stable solution found in the plot of the number of clusters. The convergence criteria or number of iteration of MRF model is the only parameter that we have varied to get multiple clusterings.

E: Segmentation of Image Using Our Proposed MSVFMM Model

Now, we put these color and texture features into our newly developed MSVFMM, where in the beginning, color and texture features are extracted from the input image using the CIE-Luv color space and Daubechies wavelet transform. Now, the data are fed into MRF model to capture the neighborhood influence among the neighboring data or pixels on the basis of these six features.

After incorporating the neighborhood relationship among the neighboring data, the K-means algorithm determines the centers of the clusters for each component of the image. Then the Expectation Maximization (EM) algorithm is used for estimating the model parameters. In fact, model parameter estimation and segmentation of pixels will run simultaneously within MAP estimation under Markovian framework. Finally, MAP will classify the pixels of the input images into different pixel classes on the basis of color and texture features. As a result, regions are obtained where there is a discontinuity in either color or texture. Here, the function Q_{MAP} plays a crucial role in determining the actual membership of a class. In fact, the color and texture features of pixel directly influence the maximization of the Q_{MAP} function. That means every individual feature has some influence on the maximization process of the Q_{MAP} function. Some features have greater influences while others have minimum effect. If the influences of some particular features are greater than that of others in the maximization process of the Q_{MAP} function for a particular pixel then the clustering of that pixel is greatly influenced by those particular features.

IV. PREVIOUS WORK

As grayscale images dominated the medical imaging domain for last three decades, segmentation approaches for color medical images are hard to find. Currently, grayscale medical imaging are gradually replacing with color medical imaging. As a result, present focus of researchers is on segmentation of color medical imaging. No significant development has been made to address this issue yet. So, we have to rely on few available grayscale medical image segmentation approaches and other color image segmentation approaches as predecessor approaches.

Several approaches have been explored by researchers for segmentation and classification of color textured patterns. Sanjay-Gopal and Herbert [1] proposed a novel approach for pixel labeling and image segmentation based on a Bayesian framework using a spatially variant finite mixture model where an EM algorithm is used for maximum likelihood estimation of the pixel labels and the parameters of the mixture densities. Dokur & Olmez [11] proposed an algorithm where feature vectors were extracted using 2D-DCT (Discrete Cosine Transform) of 4x4 pixel blocks for segmentation of ultrasonogram images by using a hybrid neural network. In addition, several similar algorithms based on DCT coefficients have been proposed by the researchers. Kato and Pong [6], proposed a segmentation technique for color textured images based on Markov Random Fields model. The proposed approach is supervised in nature. As such, no model parameter estimation technique has been used and the number of components was selected manually. In 2003, Kato, Pong, and Song [7] proposed an algorithm using a multi-layer MRF model for unsupervised segmentation that automatically estimates the number of components in the Gaussian mixture and the associated model parameters by introducing a mathematical term to the energy function of each layer. Four-paged paper does not provide enough description about how it estimates the number of components. Further, Kato and Pong [5] proposed another similar algorithm where they used color and texture feature. They utilized the EM algorithm for estimation of the model parameter but they set the number of components (pixel classes) manually. In another unsupervised segmentation approach by Panjwani & Healy [9], an Agglomerative Hierarchical Clustering for pixel segmentation has been proposed where they utilized a maximum *pseudolikelihood* scheme for model parameter estimation only. Number of components has been set manually.

Ray & Turi [12] proposed an unsupervised color image segmentation approach using K-means where they employed *Inter-Class* and *Intra-Class* distance measures for pixel classification. This is a color feature based approach where they did not consider texture features and neighborhood relationships among pixels. Yiming et al. [13] suggested another unsupervised color image segmentation approach on finite mixture model where they employed *Minimum Message Length* (MML) principle for finding the number of components in images and Maximum Likelihood Estimate (MLE) for pixel classification. This is also a color feature based approach where neither texture feature nor has neighborhood relationship been considered.

Deng & Manjunath [14] proposed another unsupervised color image segmentation approach that is mainly focused on color image segmentation using both edges and region-based approach. This approach is commonly known as JSEG approach, which has been applied on both natural color images and color medical images.

Very recently, Islam et al. [8] proposed an unsupervised image segmentation approach where they used a *Spatially Variant Finite Mixture Mode*l (SVFMM) under a multiplanner framework for color image segmentation purpose. SVFMM utilizes *Maximum a Posteriori* to segment an image into different classes or segments. Here, CIE-Luv color space and Haar wavelet transform are used to extract color and texture features respectively from an image. SNOB and *Ensembles of Clusterings* are combined to get a robust automatic tool for finding the number of components in an image.

V. ADVANTAGES OF OUR PROPOSED APPROACH OVER THE EXISTING APPROACHES

Our approach differs with other previous work in a few major ways. Approaches based on the SVFMM as proposed in other works [1], [4] have been applied on gray level images only, whereas we have developed a new multidimensional model to make it applicable for color and textured images. Further, those approaches are based on color features only. The algorithms as proposed in other previous work [5], [6], [7] are based exclusively on MRF model that uses Gabor and MRSAR texture descriptors to extract texture features. Here, the authors did not use any Finite Mixture Model in their approach rather they directly applied the MRF model in their algorithms, which makes their algorithm computationally more expensive. Further, these existing approaches are greatly influenced by similarity in color only. The proposed approaches in [9] & [11] are altogether different type of approaches in terms of clustering technique, texture feature extraction and model parameter estimation.

We have seen few unsupervised color image segmentation approaches. Among them, Kato, Pong, and Song, [7] has introduced the unsupervised technique for determining the number of pixel classes. So far, the authors have applied this approach on 3 images only out of which 2 are synthetic color images. The only natural color image they used is a very simple image that has only one object on a background. Further, we have not seen any continuation work so far on this particular approach. So, we can not consider this approach as a general approach for a wide variety of color images. Yiming et al. [13] approach is good for small applications and simple images. Deng & Manjunath's [14] JSEG approach has been found to be more effective and robust in small applications and simple natural and medical images.

In this particular work, we utilize the robust technique as used in Islam et al's work [8] to identify the number of components in an image. We also use the CIE-Luv color space for color feature extraction. They used Haar wavelet transform for texture feature extraction. We use Daubechies wavelet transform instead of Haar wavelaet transform in a bid to make the proposed approach more sensitive to texture. Further, Daubechies transforms are yet to employ as a texture descriptor under Markovian framework. The other significant difference between Islam et al's [8] approach and our proposed approach is, they utilized the SVFMM under a multiplanner framework so that it can act as a multidimensional model. Further, SVFMM was not a complete model for image segmentation. The SVFMM has been utilized for pixel classification purpose only. Other tools for preprocessing of data like color and texture descriptors and the tool for identifying the number of components were kept outside the segmentation model. In order to get more accuracy in segmentation in our proposed approach, we combine all of them in a nutshell, and name it Multidimensional Spatially Variant Finite Mixture Model (MSVFMM) that can act as a complete and independent color image segmentation tool.

VI. DISCUSSSION ON EXPERIMENTAL RESULTS

In our experiment, we have composed our images from the web site of Atlanta South Gastroenteroloy, P.C. The proposed algorithm has been tested with a variety of endoscopic images especially on stomach cancer. We have simply compared the performance of the proposed approach with a similar existing approach as proposed by Islam et al. [8] and the results are evaluated by a gastroenterologist.

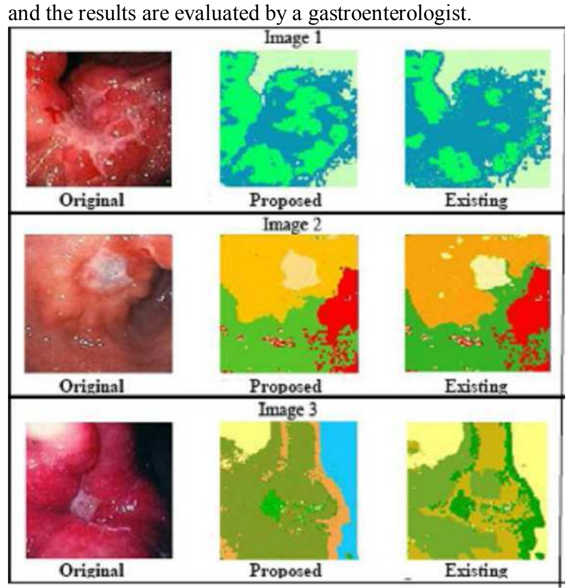


Fig. 1 Comparison of Segmentation Results.

In Fig. 1, we present 6 segmented images, 3 by using the proposed approach and the rest 3 by using similar existing approach. Our goal is to segment the ulcerated mass and adenocarcinoma lesion from the images with a greater accuracy. In image 1, there is an ulcerated mass in the middle surrounded with prominent folds of adenocarcinoma lesion.

Our proposed approach is able to differentiate the central ulcerated mass and adenocarcinoma lesion efficiently while the existing approach results in poor differentiation of ulcerated mass and adenocarcinoma. Segregated white spots are due to the usual artifacts of endoscopy system, where light reflection of the internal camera creates such white spots on the images. Similarly, into image 2, our proposed approach shows accuracy in identifying the ulcerated mass and the signet-ring type adenocarcinoma. Red region is representing the opening of esophagus in the stomach. Segregated red spots on green region at the bottom part of the image are less significant as it represents the white spot. On the other hand, the existing approach produces spurious results. It failed to identify the adenocarcinoma lesion properly as it merges lower right portion of the adenocarcinoma lesion (adjacent to the esophagus opening) into the background (green) of the image. Same things happened to image 3 as well. The existing approach segments the low intensity areas due to folds on the lesion as a separate region while the proposed approach produces an accurate segmentation where the problem of low intensity areas are successfully minimized up to an acceptable level. Better results as obtained with the proposed approach are due to its better texture discriminating ability that is what expected with the daubechies wavelet transform. Haar wavelet, on the other hand, miserably failed to discriminate textures in few areas in the images, where textures are relatively complex. Above comparison reveals that the proposed algorithm is capable to capture the finer details of the medical image in terms of color and texture using MRF model and thus able to produce accurate and faithful segmentation.

VIII. CONCLUSION

Our aim was to propose and implement a faithful tool for unsupervised segmentation of color medical images that can provide quantitative information such as size of the tumor or lesion, size of the ulceration, tissue classification, and other anatomical abnormalities. We have compared our approach with a similar alternative approach and have got better results in terms of reliability and accuracy. Future work will focus on comparison of this approach in terms of computing efficiency, robustness and segmentation accuracy with other unsupervised segmentation approaches for color textured images that can be used in medical image segmentation.

REFERENCES

[1] S. Sanjay-Gopal, and T.J. Herbert, "Bayesian Pixel Classification Using Spatially Variant Finite Mixtures and the Generalized EM Algorithm", *IEEE Trans. on Image Processing*, Vol. 7(7), July 1998, pp. 207-216.

[2] C.S. Wallace, and D.M. Boulton, "An Information measure for classification", *Computer Journal*, Vol. 11(2), pp. 185-194.

[3] C. S. Wallace, and D. L. Dow, "MML clustering of multi-state, poisson, von mises circular and gaussian distribution", *Statistics and Computing*, Vol.10(1), }', Jan. 2000, pp.73-83.

[4] K. Blekas, A. Likas, N.P. Galatsanos, and I.E. Lagaris, , "A Spatially-Constrained Mixture Model for Image Segmentation", *IEEE Trans. on Neural Networks*, Vol. 16(2), pp. 494-498.

[5] Z. Kato, and T.C. Pong, "A Markov Random Field Image Segmentation Model for Color Textured Images", *Image and Vision Computing*, Vol. 24(10), pp. 1103-1114.

[6] Z. Kato, and T. C. Pong, "A Markov Random Field Image Segmentation Model for Combined Color and Textured Features, W.Skarbek (ED)", *Proceedings of ICCAIP*, Vol. 2124 of LNCS, Springer, Warsaw', pp. 547-554.

[7] Z. Kato, T.C. Pong, and G. Song, "Unsupervised segmentation of color textured images using a multi-layer MRF model", *Proceedings of ICIP*, Vol. (1), Spain, (2003) pp. 961-964.

[8] M. Islam, J. Yearwood, and P. Vamplew, "Unsupervised Color Textured Image Segmentation Using Cluster Ensembles and MRF Model", Tarek Sobh (ED)", *Proceedings of CISSE 2007*, Vol. Advances in Computer and Information Sciences and Engineering, Springer-Verlag', pp. 323-328.

[9] D. K. Panjwani, and G. Healy, "Markov Random Field Models for Unsupervised Segmentation of Textured Color Images", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 17(10).

[10] G.R. Shivani, P. Manika, and D. Shukhendu, "Unsupervised Segmentation of Texture Images Using a Combination of Gabor and Wavelet Features", ICVGIP 2004, *Proceedings of the Fourth Indian Conference on Computer Vision, Graphics & Image Processing*, Kolkata, India', Vol. December 16-18, 2004}, pp. 370-375.

[11] Dokur, Z., and Olmez, T.:Segmentation of Ultrasound images by Using a Hybrid Neural Network, Pattern Recognition Letters, Vol. 23 (2002) 1825-1836.

[12] R. Siddheswar and R.H. Turi, "Determination of Number of Clusters in k-means Clustering and application in Color Image Segmentation", *In Proceedings of the 4th Intl. Conf. on Advances in Pattern Recognition and Digital Techniques* (ICAPRDT'99), vol. Calcutta, India, 1999 pages: 137-143.

[13] Wu Yiming, Yang Xiangyu, and Chan Kap Luk, "Unsupervised Color Image Segmentation based on Gaussian Mixture Model", In *Proceedings of the 2003 Joint Conf. of the 4th Intl. Conf. on Information, Communications and Signal Processing*, Vol. 1(15-18 Dec. 2003), pages: 541-544.

[14] Y. Deng, and B.S. Manjunath, "Unsupervised Segmentation of Color-Texture Regions in Images and Videos", In *IEEE Trans. on Pattern.*

[15] Randen and J.H. Husoy, "Filtering for Texture Classification: A Comparative Study", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. April, 1999, pp. 291-310.*Analysis and Machine Intelligences*, Vol. 23(8), pp. 800-810.

[16] Unser, M.:Texture classification and Segmentation Using Wavelet Frames, IEEE Trans. on Image Processing, Vol. 4, No. 11, (1995) 1549-1560.

MRF Model Based Unsupervised Color Textured Image Segmentation Using Multidimensional Spatially Variant Finite Mixture Model

Mofakharul Islam Peter Vamplew John Yearwood
Center of Informatics and Applied Optimization
School of Information Technology and Mathematical Sciences
University of Ballarat, Ballarat,
Australia

Abstract- **We investigate and propose a novel approach to implement an unsupervised color image segmentation model that segments a color image meaningfully and partitions into its constituent parts automatically. The aim is to devise a robust unsupervised segmentation approach that can segment a color textured image more accurately. Here, color and texture information of each individual pixel along with the spatial relationship within its neighborhood have been considered for producing more accuracy in segmentation. In this particular work, the problem we want to investigate is to implement a robust unsupervised Multidimensional Spatially Variant Finite Mixture Model (MSVFMM) based color image segmentation approach using Cluster Ensembles and MRF model along with Daubechies wavelet transforms for increasing the content sensitivity of the segmentation model in order to get a better accuracy in segmentation. Here, Cluster Ensemble has been utilized as a robust automatic tool for finding the number of components in an image. The main idea behind this work is introducing a Bayesian inference based approach to estimate the Maximum a Posteriori (MAP) to identify the different objects/components in a color image. Markov Random Field (MRF) plays a crucial role in capturing the relationships among the neighboring pixels. An Expectation Maximization (EM) model fitting MAP algorithm segments the image utilizing the pixel's color and texture features and the captured neighborhood relationships among them. The algorithm simultaneously calculates the model parameters and segments the pixels iteratively in an interleaved manner. Finally, it converges to a solution where the model parameters and pixel labels are stabilized within a specified criterion. Finally, we have compared our results with another recent segmentation approach [10], which is similar in nature. The experimental results reveal that the proposed approach is capable of producing more accurate and faithful segmentation and can be employed in different practical image content understanding applications.**

I. INTRODUCTION

Image Segmentation is a process that isolates homogenous regions of images in a meaningful manner. Using a finite mixture model in color image segmentation is a popular approach as in statistical model-based methods, computationally intensive algorithms are a common drawback. As a result, finite mixture model has received significant attention from researchers in last decade. Finite mixture models have been further modified by Sanjay-Gopal and Herbert [1] to Spatially Variant Finite Mixture Model (SVFMM) by introducing a process based on Markov Random Fields (MRF) model, which is able to capture the spatial relationship among the neighboring pixels. So, SVFMM has some advantageous features, especially in the case of contextual-dependent physical phenomena like images as it incorporates the outstanding features of MRF model without additional computational burden. In the SVFMM method, MRF introduces a prior distribution that takes into account the neighborhood dependency or relationship among the neighboring pixels. Recently, approaches using MRF models and Bayesian methods have provided answers to various contemporary problems in image segmentation [1], [4], [5], [6], [7], [8], [9], [10] and [11]. The theoretical framework relies on Bayesian estimation via combinatorial optimization. Different pixels represent different segmented regions in the input image. These classes are represented by multivariate Gaussian distribution. We elaborate a stochastic model based segmentation framework based on Gaussian Mixture Models. In this paper, we use CIE-Luv color space as color features and a set of texture features extracted through Daubechies wavelet transform.

We propose a Gaussian parameter estimation method using the EM-algorithm. The application of a clustering method to image segmentation has the particular characteristics that spatial information should be taken into account for segmentation accuracy. Here in addition to intensity-values of pixels, the pixel location must also be used to determine the cluster to which each pixel is assigned. Conceptually, in most cases it is desired that the same cluster label be assigned to spatially adjacent pixels. In implementing these ideas, the Bayesian framework provides a natural approach. Pixel intensity information is captured by the likelihood term in Bayesian Networks, while a prior biasing term captures the spatial location information with the help of MRF model.

Due to the complexity of determining color distance in RGB spaces, several simple formulas have been proposed, which approximate the Reimannian space by a Euclidean color space yielding perceptually uniform color spacing. One of them is the CIE-Luv color model that we use in our present work.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_68,

Considering the advantages and disadvantages of different texture descriptors, we have chosen transform domain features which refers to a mathematical representation of an image. The main transform domain features used in the past are: Discrete Fourier Transform (DFT) and Discrete wavelet Transform (DWT). Fourier analysis consists of breaking up a signal into sine waves of various frequencies, whereas wavelet analysis breaks up a signal into shifted and scaled versions of an original wavelet that refers to decomposition of a signal with a family of basis functions obtained through translation and dilation of a special function [15]. Moments of wavelet coefficients in various frequency bands have been shown to be effective for representing subtle variation in texture [16]. Thus we prefer to use wavelet transform to extract texture features in our work.

Recently, researchers have found that segmentation based purely on texture gives fuzzy boundaries, but usually homogenous regions. On the other hand, segmentation based on color is more sensitive to local variations, but provides sharp boundaries. Thus by taking into account combined features (color and texture), the advantage of both color and texture has been preserved, as a result of which sharp boundaries and homogenous regions are obtained [5].

The remainder of this paper proceeds as follows: In Section II we describe the proposed Multidimensional Spatially Variant Finite Mixture Model (MSVFMM) that we will use in our proposed approach for image segmentation. Section III presents a novel MSVFMM based approach for color image segmentation. In Section IV, we present previous work in this specific field followed by Section V, where new features of the proposed approach are discussed. Experimental results demonstrating the accuracy and efficiency of the proposed approach are discussed in Section VI and finally in Section VII, we present our conclusion and future work.

II. MULTIDIMENSIONAL SPATIALLY VARIANT FINITE MIXTURE MODEL (MSVFMM)

From an image segmentation point of view, approaches based on the SVFMM as proposed in other works [1], [4] have a common limitation that it can handle single feature only. The reason is, the SVFMM is a single dimensional model. As a result, its application is limited to gray level images only. In our proposed approach, we have improved and modified the SVFMM into a full functional MSVFFM to make it efficient for longer set of features, which makes it enable to handle multiple feature set. Therefore, our proposed multidimensional SVFMM is capable to handle multiple cues including color and texture features simultaneously and efficiently. Further, we have integrated the number of components finding tool and features extraction tools into the model to make it as a robust color image segmentation approach that eventually can handle a wide variety of natural and/or industrial color images.

Basically, SVFM model is a classical finite mixture model that employs MRF model to capture the neighborhood relationships among the neighboring pixels. We have proposed a *multidimensional spatially variant finite mixture model* (MSVFMM) which is able to handle a large feature vector rather than a single feature as proposed by Sanjay-Gopal & Herbert and Blekas et al., which are in fact, single dimensional [1], [4]. In statistical model-based methods, computationally intensive algorithms are a common drawback. So, finite mixture model in image segmentation is a popular approach. As a result, classical finite mixture models have received significant attention from researchers in the last decade. Finite mixture model has been further modified by Sanjay-Gopal and Herbert [1] and Blekas et al. [4] to *spatially variant finite mixture models* (SVFMM) by introducing a process based on Markov Random Fields (MRF), which is able to capture the spatial relationship among neighboring pixels. In the proposed MSVFMM method, MRF introduces a prior distribution that takes into account the neighborhood dependency or relationship among the neighboring pixels. We elaborate a stochastic model based segmentation framework based on Gaussian Mixture Models, which constitute a well-known probabilistic Neural Network Model.

In fact, the SVFMM model is a modification of the classical mixture model or Gaussian mixture model with the introduction of MRF model. We assume a mixture model with an unknown number of components K each one having its own vector of density parameter θ_j for each feature.

In our multidimensional model, each ith pixel of the image is represented as a vector of values as $\chi^i = (x_1^i, x_2^i, \ldots., x_n^i)$ where x_n^i denotes pixel values in the nth dimensional plane. So, the probabilities of the ith pixel belonging to the jth class label are given as;

$$\eta_j^i = P(j \mid x^i) \tag{1}$$

where, x^i = observation at the ith pixel. Here, model parameters should strictly satisfy several constraints at the end of this section.

Similar to the SVFMM, we assume that the density function $f(x_i \mid \Phi, \Gamma)$ at any observation x_i is given by:

$$f(x_i \mid \Phi, \Gamma) = \sum_{j=1}^{K} p_j^i \psi(x_i \mid \theta^j) \tag{2}$$

where, $\psi(x_i \mid \theta^j)$ is a Gaussian distribution with parameters $\theta_j = (\mu_j, \sigma_j)$, and K is the number of components in the mixture.

Maximum a posteriori (MAP) introduces a prior distribution for the parameter set Φ that takes into account spatial information based on the *Gibbs function*. According to the Hammersley-Clifford theorem, *Gibbs distribution* takes the following form;

$$P(\Phi) = \tfrac{1}{Z} \exp(-U(\Phi)) \tag{3}$$

where, $U(\Phi) = \beta \sum V_{N_i}(\Phi)$.

Φ is a vector of features. The function $-U(\Phi)$ is an energy term. β is called the *regularization parameter*, the normalizing constant Z is called a partition function and $V_{N_i}(\Phi)$ denotes *clique potential* of the label configuration p^m within the neighborhood N_i of the *i*th pixel which can be calculated as;

$$V_{N_i}(\Phi) = \sum_{m \in N_i} g(u_{i,m}) \tag{4}$$

Where the $u_{i,m}$ denotes the distance between the two label vectors p^i and p^m. The function *g(u)* must be gradually increasing and nonnegative.

A posterior log density function can be derived from (3) as;

$$P(\Phi, \Gamma \mid X) = \sum_{i=}^{N} \log(x^i \mid \Phi, \Gamma) + \log P(\Phi) \tag{5}$$

The *Expectation-Maximization* (EM) algorithm requires that the computation of the conditional expectation value z_j^i of the hidden variables must be at the *Expectation* step due to MAP estimation of the parameters $\{p_j^i\}$ and $\{\theta_j\}$;

$$z_j^{i^{(t)}} = \frac{p_j^{i^{(t)}} \psi(x^i | \theta_j^{(t)})}{\sum_{l=1}^{K} p_l^{i^{(t)}} \psi(\chi^i | \theta_j^{(t)})} \tag{6}$$

Then, maximization of the following log-likelihood corresponding to the complete data set is performed in the *Maximization* step;

$$Q_{MAP}(\Phi, \Gamma \mid \Phi^{(t)} \Gamma^{(t)}) \tag{7}$$

where, t = iteration step.

For each parameter, Q_{MAP} can be maximized independently, which yields the update equations for parameter of the component densities $\mu_j^{(t+1)}$ and $[\sigma_j^2]^{(t+1)}$. However, maximization of the function Q_{MAP} with respect to the label parameter p_j^i does not provide a closed form of update equations. In addition, the maximization procedure must also take into the constraints $0 \le p_j^i \le 1$ and $\sum_{j=1}^{K} p_j^i = 1$. However, this difficulty has been encountered successfully by Sanjay-Gopal & Herbert [1] and subsequently by Blekas et al [4]. Sanjay-Gopal and Herbert [1] has introduced a Gradient Projection (GP) algorithm to make the EM algorithm using closed form expressions while Blekas et al. [4] implemented the M-step based on a closed form update equation followed by an efficient projection method.

III. PROPOSED COLOR IMAGE SEGMENTATION APPROACH

In different imaging applications like medical imaging, biometrics, security and surveillance, mineral and mining, and material science where the primary goal is accuracy in segmentation, subtle information related to color and texture can play significant role in achieving this goal. In this particular work, we have used CIE-Luv and Daubechies wavelet transforms as color and texture descriptors respectively. We combine the CIE-Luv color space descriptor, Daubechies transforms wavelet descriptors, a robust number of components finding tool using SNOB and Cluster Ensembles, and SVFMM model into an integrated segmentation model (MSVFMM), which is able to segment a wide variety of color images meaningfully with a greater accuracy.

The segmentation quality greatly depends on the choice of feature (intensity, color, texture, and co-ordinate) descriptor. So, feature extraction is the most significant aspect in image segmentation.

Choosing of a suitable color space is very important as the quality or accuracy in segmentation largely depends on it. Most of the color spaces are three-dimensional where different dimensions represent the different color components. The most common color space in computer graphics is RGB. Unfortunately, this 3D color space does not correspond to equal perception of color dissimilarity thus not suitable for human perception. As a result, alternative color spaces like HSV, CIE-Lab and CIE-Luv are generated by nonlinear transformation of the RGB color space. We prefer CIE-Luv as it represents the three characteristics i.e., hue, lightness and saturation that characterize the perceptually uniform color efficiently.

The texture descriptor plays a significant role in the analysis of contextually dependent physical phenomena like image in addition to the color descriptor. This project aims to investigate and implement a more context sensitive image segmentation technique that can produce reliable and accurate segmentation of color images based on subtle color and texture variation. We use wavelet transform to extract texture features. Although, wavelet transform has made significant contribution in the several areas of image processing such as image enhancement, image compression, and image registration, its use in image segmentation is very limited. Using Wavelet transform is a recent trend in image processing. Very recently, Islam et al. [10] used Haar wavelet transform to extract texture features from natural color images in their image segmentation approach and found promising results. In this particular project, we use Daubechies wavelet transform, which is found to be more accurate and sensitive in capturing texture details of a color image. We have introduced Daubechies wavelet transforms for the first time as texture descriptor under Markovian framework for color image segmentation purpose. Image analysis in different imaging application domains demands a more sensitive texture descriptor that can provide us an efficient way to capture the subtle variation in texture. The Daubechies wavelet transforms has got the ability to capture texture information more precisely and accurately while other texture descriptors are not ideal in those specific situations.

Recently, approaches using MRF models and Bayesian methods have provided answers to many contemporary problems in image segmentation. The application of a clustering method to image segmentation has the particular characteristics that spatial information should be taken into consideration. In addition to intensity-values of pixels, the pixel location must also be used to determine the cluster to which each pixel is assigned. Conceptually, in most cases it is desired that the same cluster label be assigned to spatially adjacent pixels. In implementing these ideas, the Bayesian framework provides a natural approach. Pixel intensity information is captured by the likelihood term in Bayesian Networks, while a prior biasing term captures the spatial location information with the help of Markov Random Fields (MRF) model.

A. Extraction of Color Feature

Color is an important dimension of human visual perception that allows discrimination and recognition of visual information. Extracting and matching color features are relatively easy. In addition to that, color features have been found to be effective in color image segmentation [5], [6], [7]. In this particular work, we prefer CIE-Luv as it represents the three characteristics i.e., hue, lightness and saturation that characterize the perceptually uniform color efficiently. In addition, CIE-Luv is preferred by the computer vision and graphics industry.

B. Extraction of Texture Feature

Texture characterizes local variations of pixel's color or intensity. There is no formal or unique definition of texture though texture based methods are commonly being used in computer vision and graphics. Each texture analysis method defines texture from its own perspective. Texture can be defined as a local statistical pattern of texture primitives in observer's domain of interest. Texture is often describe as consisting of primitives that are arranged according to placement rule which gives rise to structural methods of texture analysis that explicitly attempts to recover the primitives and the replacement rules. On the other hand, statistical methods consider texture as random phenomena with identifiable local statistics. The neighborhood property is a common similarity among all the texture descriptions. So, we can define texture as an image feature which is characterized by the gray value or color pattern in a neighborhood surrounding the pixel.

We can use many texture extraction tools or descriptors to extract texture feature. Using Wavelet transform is a recent trend in image processing. Although other texture descriptors like Gabor and MRSAR filters [9], [17] are popular but from the computational burden point of view, wavelet transform performs better than other texture descriptors. Gabor filters are found efficient in capturing the strong ordered or coarser texture information while MRSAR filters works well in capturing weak ordered or finer texture details. Wavelet transform, on the other hand, is found to be efficient in capturing both the coarser and finer texture information simultaneously with lesser computational burden.

Having been inspired by Islam et al's [10] work, where they used Haar wavelet transform and found promising results in capturing texture information, we have decided to utilize Daubechies wavelet transform, which is more accurate in discriminating texture. The Daubechies wavelet transform is named after its inventor, the mathematician Ingrid Daubechies. The Daubechies D4 transform has four wavelet and scaling function coefficients.

C. Finding The Number of Components

We have tried several data clustering approaches to identify the number of components automatically but most of them have been found inappropriate for handling image data properly. A few of them are found incapable to handle data having corrupted values as many pixels may not have true values for some attributes while few of them are not performing well in handling discrete data like image data. SNOB [2], [3], a *Minimum Message Length* (MML) based unsupervised data clustering approach is free from this sort of difficulties and has been found more effective in handling noisy color image data that have a diverse varieties of faults and imperfections.

In general SNOB has been found more or less effective in identifying the number of components in an image but sometimes it produces spurious results due to its limitations to handle some imaging artifacts like noise, shadow, and intensity inhomogeneity. In order to remove this difficulty of SNOB, we utilize the technique *Ensembles of Clusterings*. We use SNOB as a starting point for getting multiple clusterings based on which Ensembles of Clusterings produce a consolidated clustering. The procedure we have followed for determining the number of components is as follows: produce multiple clusterings for different parameter values and select the stable solution found in the plot of the number of clusters. The convergence criteria or number of iteration of MRF model is the only parameter that we have varied to get multiple clusterings.

E: Segmentation of Image Using the Newly Developed MSVFMM Model

Now, we put these color and texture features into our newly developed MSVFMM, where in the beginning, color and texture features are extracted from the input image using the CIE-Luv color space and Dabechies wavelet transform. Now, the data are fed into MRF model to capture the neighborhood influence among the neighboring data or pixels on the basis of these six features.

After incorporating the neighborhood relationship among the neighboring data, the K-means algorithm determines the centers of the clusters for each component of the image. Then the Expectation Maximization (EM) algorithm is used for estimating the model parameters. In fact, model parameter estimation and segmentation of pixels will run simultaneously within MAP estimation under Markovian framework. Finally, MAP will classify the pixels of the input images into different pixel classes on the basis of color and texture features. As a result, regions are obtained where there is a discontinuity in either color or texture. Here, the function Q_{MAP} plays a crucial role in determining the actual membership of a class. In fact, the color and texture features of pixel directly influence the maximization of the Q_{MAP} function. That means every individual feature has some influence on the maximization process of the Q_{MAP} function. Some features have greater influences while others have minimum effect. If the influences of some particular features are greater than that of others in the maximization process of the Q_{MAP} function for a particular pixel then the clustering of that pixel is greatly influenced by those particular features.

IV. RELATED PREVIOUS WORK

Kato & Pong proposed a segmentation technique for color textured images based on Markov Random Fields using color and texture features [6]. The proposed approach is supervised in nature. As such, no model parameter estimation technique has been used and the number of components was selected manually. In 2003, Kato, Pong, and Song proposed another color and texture feature based algorithm using a multi-layer MRF model for unsupervised segmentation that automatically estimates the number of components in the Gaussian mixture and the associated model parameters by introducing a mathematical term to the energy function of each layer [7]. Four-paged paper does not provide enough description about how it estimates the number of components. Further, Kato & Pong proposed another similar algorithm where they used color and texture feature. They utilized the EM algorithm for estimation of the model parameter but they set the number of components (pixel classes) manually [5]. Panjwani & Healy proposed another unsupervised segmentation approach, where *Agglomerative Hierarchical Clustering* has been utilized for pixel segmentation and a maximum pseudolikelihood scheme for model parameter estimation [8]. No technique has been applied for determination of number of components.

Ray & Turi [11] proposed an unsupervised color image segmentation approach using K-means where they employed *Inter-Class* and *Intra-Class* distance measures for pixel classification. This is a color feature based approach where they did not consider texture features and neighborhood relationships among pixels. Yiming et al. [12] suggested another unsupervised color image segmentation approach on finite mixture model where they employed *Minimum Message Length* (MML) principle for finding the number of components in images and Maximum Likelihood Estimate (MLE) for pixel classification. This is also a color feature based approach where neither texture feature nor has neighborhood relationship been considered.

Deng & Manjunath [13] proposed another unsupervised color image segmentation approach that is mainly focused on color image segmentation using both edges and region-based approach. This approach is commonly known as JSEG approach.

Very recently, Islam et al. [10] proposed an unsupervised image segmentation approach where they used a *Spatially Variant Finite Mixture Mode*l (SVFMM) under a multiplanner framework for color image segmentation purpose. SVFMM utilizes *Maximum a Posteriori* to segment an image into different classes or segments. Here, CIE-Luv color space and Haar wavelet transform are used to extract color and texture features respectively from an image. SNOB and *Ensembles of Clusterings* are combined to get a robust automatic tool for finding the number of components in an image.

V. WHAT'S NEW IN OUR PROPOSED APPROACH

Our approach is different from previous work in a few major ways. Approaches based on the SVFMM as proposed in other works [1], [4] have been applied on gray level images only, whereas we have developed a new multidimensional model to make it applicable for color and textured images. Further, those approaches are based on color features only. The algorithms as proposed in other previous work [5], [6], [7] are based exclusively on MRF model that uses Gabor and MRSAR texture descriptors to extract texture features. Here, the authors did not use any Finite Mixture Model in their approach rather they directly applied the MRF model in their algorithms, which makes their algorithm computationally more expensive. Further, these existing approaches are greatly influenced by similarity in color only.

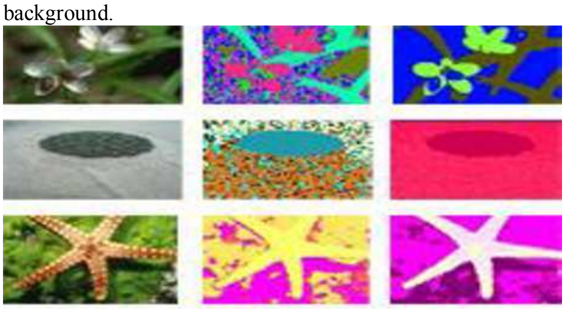
We have seen few unsupervised color image segmentation approaches. Among them, Kato, Pong, and Song, [7] has introduced the unsupervised technique for determining the number of pixel classes. So far, the authors have applied this approach on 3 images only out of which 2 are synthetic color images. The only natural color image they used is a very simple image that has only one object on a background. Further, we have not seen any continuation work so far on this particular approach. So, we can not consider this approach as a general approach for a wide variety of color images. Yiming et al. [12] approach is good for small applications and simple images. Deng & Manjunath's [13] JSEG approach has been found to be more effective and robust in small applications and simple images.

In this particular work, we utilize the robust technique as used in Islam et al's work [10] to identify the number of components in an image. We also use the CIE-Luv color space for color feature extraction. They used Haar wavelet transform for texture feature extraction. We use Daubechies wavelet transform instead of Haar wavelaet transform in a bid to make the proposed approach more sensitive to texture. Further, Daubechies transforms are yet to employ as a texture descriptor under Markovian framework. The other significant difference between Islam et al's [10] approach and our proposed approach is, they utilized the SVFMM under a multiplanner framework so that it can act as a multidimensional model. Further, SVFMM was not a complete model for image segmentation. The SVFMM has been utilized for pixel classification purpose only. Other tools for preprocessing of data like color and texture descriptors and the tool for identifying the number of components were kept outside the segmentation model. In our proposed approach, we combine all of them in a nutshell, and name it *Multidimensional Spatially Variant Finite Mixture Model* (MSVFMM) that can act as a complete and independent color image segmentation tool.

VI. DISCUSSSION ON EXPERIMENTAL RESULTS

In our experiment, we have sourced our images from the Barkeley Segmentation dataset [14]. We have compared our proposed approach with another similar color image segmentation approach proposed by Islam et al. [10]. Our approach has been tested on a wide variety of natural images and found promising results but we could not present these results here due to space constraint.

In Fig. 1, a comparison has been shown. In the top image, it is clearly evident that there are some sorts of over segmentation in the result as obtained from the existing approach while on the other hand, the proposed approach has done the job perfectly. In the middle image, again we have experienced over segmentation with the existing approach. It failed to segment properly the surrounding portion in the background.



Original Existing Proposed

Fig. 1. Comparison of Segmentation Results.

In the third image, results show that the proposed approach clearly outperformed the existing approach in terms of segmentation accuracy. Better results as obtained with the proposed approach are due to its better texture discriminating ability that is what expected with the daubechies wavelet transform. Haar wavelet, on the other hand, miserably failed to discriminate textures in few areas in the images, where textures are relatively complex. It is clearly evident from the comparison of results that our proposed approach is able to segment an image into its constituent objects in a more meaningful manner than the existing approach.

VII. CONCLUSION

We have proposed a novel MSVFMM based approach that employs Daubechies wavelet transform for texture feature extraction, which eventually makes the approach more sensitive to the texture and thus able to perform better than the existing approach that employs Haar wavelet transform in SVFMM model. The proposed approach can successfully segment relatively complex color images with greater accuracy. Future work will focus on comparison of this approach with similar MRF based approaches, where other color and texture descriptors (especially Gabor and MRSAR) have been used.

REFERENCES

[1] S. Sanjay-Gopal, and T.J. Herbert, "Bayesian Pixel Classification Using Spatially Variant Finite Mixtures and the Generalized EM Algorithm", *IEEE Trans. on Image Processing*, Vol. 7(7), July 1998, pp. 207-216.

[2] C.S. Wallace, and D.M. Boulton, "An Information measure for classification", *Computer Journal*, Vol. 11(2), pp. 185-194.

[3] C.S. Wallace, and D..L. Dow, "MML clustering of multi-state, poisson, von mises circular and gaussian distribution", *Statistics and Computing*, Vol.10(1), }', Jan. 2000, pp.73-83.

[4] K. Blekas, A. Likas, N.P. Galatsanos, and I.E. Lagaris, , "A Spatially-Constrained Mixture Model for Image Segmentation", *IEEE Trans. on Neural Networks*, Vol. 16(2), pp. 494-498.

[5] Z. Kato, and T.C. Pong, "A Markov Random Field Image Segmentation Model for Color Textured Images", *Image and Vision Computing*, Vol. 24(10), pp. 1103-1114.

[6] Z. Kato, and T. C. Pong, "A Markov Random Field Image Segmentation Model for Combined Color and Textured Features, W.Skarbek (ED)", *Proceedings of ICCAIP*, Vol. 2124 of LNCS, Springer, Warsaw', pp. 547-554.

[7] Z. Kato, T.C. Pong, and G. Song, "Unsupervised segmentation of color textured images using a multi-layer MRF model", *Proceedings of ICIP*, Vol. (1), Spain, (2003) pp. 961-964.

[8] D.K. Panjwani, and G. Healy, "Markov Random Field Models for Unsupervised Segmentation of Textured Color Images", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 17(10).

[9] G.R. Shivani, P. Manika, and D. Shukhendu, "Unsupervised Segmentation of Texture Images Using a Combination of Gabor and Wavelet Features", ICVGIP 2004, *Proceedings of the Fourth Indian Conference on Computer Vision, Graphics & Image Processing*, Kolkata, India', Vol. December 16-18, 2004}, pp. 370-375.

[10] M. Islam, J. Yearwood, and P. Vamplew, "Unsupervised Color Textured Image Segmentation Using Cluster Ensembles and MRF Model", Tarek Sobh (ED)", *Proceedings of CISSE 2007*, Vol. Advances in Computer and Information Sciences and Engineering, Springer-Verlag', pp. 323-328.

[11] R. Siddheswar and R.H. Turi, "Determination of Number of Clusters in k-means Clustering and application in Color Image Segmentation", *In Proceedings of the 4th Intl. Conf. on Advances in Pattern Recognition and Digital Techniques* (ICAPRDT'99), vol. Calcutta, India, 1999 pages: 137-143.

[12] Wu Yiming, Yang Xiangyu, and Chan Kap Luk, "Unsupervised Color Image Segmentation based on Gaussian Mixture Model", In *Proceedings of the 2003 Joint Conf. of the 4th Intl. Conf. on Information, Communications and Signal Processing*, Vol. 1(15-18 Dec. 2003), pages: 541-544.

[13] Y. Deng, and B.S. Manjunath, "Unsupervised Segmentation of Color-Texture Regions in Images and Videos" In *IEEE Trans. on Pattern*.

[14] The Berkeley Segmentation Dataset (http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/BSDS300/html.

[15] Manjunath, B. S., Ma,W. Y. *"Texture Features for Image Retrieval"* John Wily & Sons, New York, 2002, pp. 313-344.

[16] Unser, M. "Classification and Segmentation Using Wavelet Frames" IEEE Trans. Image Processing, Vol. 4, No. 11, pp. 1549-1560,. 1995.

[17] Randen and J.H. Husoy, "Filtering for Texture Classification: A Comparative Study", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. April, 1999, pp. 291-310.*Analysis and Machine Intelligences*, Vol. 23(8), pp. 800-810.

Creating a Novel Multi Platform Technology Based System for School Quality Assessment

Anirudh Roy Popli, Preeti Choudhary
Indian Institute of Management Bangalore, India
{anirudhr07; preetic07}@iimb.ernet.in

Abstract – **Education in the recent years has received a lot of attention both from the academicians as well as the policy makers. Specifically in India a lot of new initiatives have been implemented but unfortunately the results have not been very encouraging. This has necessitated the presence of some sort of quality assessment system at the grass-root level. In this paper we have looked at one such recent initiative by the Karnataka State Government. We have further critically analyzed the working of this organization comparing it with the similar initiatives in the developed countries. Finally we have proposed a novel technology based assessment system which integrates different information flows into an overarching central information reservoir or information system. The ultimate motive is to mitigate the inefficiencies and deficiencies of the existing system and in the process create a system which ensures learning and development of students – the oft forgotten primary beneficiaries.**

I. Introduction

As India transitions into a knowledge economy, education both at the primary and at the higher level is receiving increasing attention from academicians as well as the policy makers. Not only various new initiatives being designed and implemented but also a dedicated resource base is being allocated to the education sector. The planned spend on the education sector by the government has grown five folds in less than a decade. The ingenuity of the policy makers is not restricted to the new initiatives like Sarva Shiksha Abhiyan and Mid Day Meal Program but rather extends to the tax regulations which require the Indian citizens to pay an education cess used to fund various new educational schemes. But unfortunately these initiatives have not been very successful in terms of meeting the expectations of the policy makers or the common man in India. This has necessitated the presence of a quality assessment system which would not only provide a check for the quality of education being imparted to the students but also provide a measure for evaluating the success of the various new initiatives.

In this paper we present the case of KSQAO (Karnataka School Quality Assessment Organization) a novel initiative by the Karnataka Government. We further refer to a few examples of how quality assessment in education is happening across the world and what are some of the best practices. Finally we develop and present a novel technology based solution that ensures removal of various inefficiencies in the existing system with the ultimate aim of providing efficient learning opportunities to the students – the eventual beneficiaries.

II. KSQAO – Case study

A. *Karnataka School Quality Assessment Organization (KSQAO)*

Karnataka School Quality Assessment Organization was established by the government of Karnataka in May 2005. Set up under the purview of department of education, its primary aim was to assess the quality of education being imparted in the schools of Karnataka. The definition of quality is not restricted to the level of learning of the students and includes additional parameters like enrollment of eligible students, their attendance and long term retention etc., parameters that are considered crucial for a successful educational initiative. The objectives established for KSQAO although many, in essence require it to assess the learning outcome of the students and provide the various reports and analysis to the:

1.) Parents and the common man at large to create awareness regarding importance of quality education for the society.
2.) Planners and policy makers, to help them design the right initiatives to aid continuous enhancement of quality of education.
3.) Schools and various functionaries in order to promote self analysis and transformation of the class room environment to aid enhanced learning for the students [4].

To achieve these objectives KSQAO has not only developed a comprehensive assessment process, it has also taken various initiatives to motivate the teachers and various nodal functionaries through training programs, awareness workshops, providing course material etc. The actual assessment process consists of the following phases:

1.) Program Development: An assessment program is developed by the state core committee. This includes taking decisions like which schools, grades etc. would be included in the assessment and what modes of assessment would be employed etc.
2.) Question Paper Development: Once the strategic decisions have been taken, a team of teachers and District Institute of Education and Training (DIET) scholars is put into place for creating questions for the assessment. A review committee then reviews and finalizes the question papers. Pilot

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_69,

tests are conducted in a few schools to check the veracity of these papers which undergo another churn before being finalized for the state wide assessment.

3.) Administration of the Assessment: Task forces constituting DIET scholars, teachers and principals along with volunteers from philanthropic organizations like Akshara Foundation and Azim Premji Foundation are created at the block and cluster level.

4.) Evaluation of the Answers: A team of four evaluators is assigned to each school which is responsible for evaluating and grading the answer sheets. These evaluated sheets are then scanned to save the information in the digital form on the central servers.

5.) Analysis and report generation: Compiled data is then analyzed in order to understand the performance at different levels e.g. state, district, cluster and school level. School and cluster level report cards are generated and sent to the schools while a state level report is sent to the policy makers and uploaded on the KSQAO website for general viewing [4].

TABLE I
PERCENTAGE OF STUDENTS WHO CLEARED THE ASSESSMENT TESTS

Years/Grades	Class 2/3	Class 5	Class 7
2005-06	67	49	48
2006-07	73.6	59.6	62.7
2007-08	-	70.8	71.28

B. Areas of Concern

Although the whole setup appears to be well thought out once we started looking at it in more detail we came across a few areas of concern:

1.) *Actions are not in-line with the Objectives:* KSQAO has recently undertaken the task of distributing teachers with activity guide books to help them prepare their students for KSQAO. I feel that this defeats the very purpose of the assessment. The end objective is to provide the policy makers with a realistic picture of the outcomes of their past initiatives and if possible provide them with an insight into the problems at the grass-root levels to help them develop better initiatives. Even otherwise if the objective is to improve the quality of learning of a student, a better solution would be to help teachers improve the classroom activities and provide them with regular feedback on the learning outcomes of students rather than to coach them for just one specific exam like KSQAO.

2.) *Assessment is conducted by people affected by its outcome – Teachers:* Although KSQAO has tried its best to make teachers aware of the fact that this assessment does not have any bearing on their job or career prospects, but the teachers remain skeptical. And their reasons are quite valid. Even though KSQAO does not take any action good or bad, against any teacher based its assessment, the assessment results are used by the officials of the education department to compare schools. Schools and sometimes teachers are asked to justify their poor performance. This creates a lot of pressure on the schools and motivates them to adopt malpractices (despite shuffling of teachers) to ensure that their students perform well in the KSQAO assessment.

3.) Assessment process is vulnerable to malpractices: Although some thought has been given to build a robust assessment process, this objective is far from being achieved. There have been instances when students of a higher grade are made to write the test for the lower grade's assessment or teachers helping the students during the test or easiest of all teachers allowing students to cheat during the assessment. Conducting the assessment on a pre announced date and using the same set of question papers across different phases (phases are inevitable given the sheer size of the state) rather add to the malpractices woes of KSQAO. Even the evaluation is manual leaving a lot of room for human error either intentional or unintentional.

III. EXISTING BEST PRACTICES

Given that KSQAO is still young we felt that a good starting point for our work and subsequent recommendations would be to look at the existing best practices across the world, especially the developed countries like US and UK were a lot work is happening in this field. Further our research showed that work in this area had been prominently in two areas, one pertaining to the definition and measurement of quality as applied to education and the other pertaining to the actual assessment and its relevance to learning - the end objective. In the subsequent paragraphs we discuss some examples and best practices which we felt might be useful to our discussion.

1.) Bloom's Taxonomy: Although the efficacy of Bloom's taxonomy is much debated we found that it is still one of the most well established forms of assessment of learning. The taxonomy provides a hierarchically inclusive form of assessment where one level subsumes all the levels below it. These six levels in ascending order are Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. The implication for our work is to ensure that the assessment does not restrict itself to evaluating rote learning. Only when comprehension is complemented by application and thinking we can assume learning. This kind of evaluation also provides significant insight into the areas that require strengthening when it comes to student's learning.

2.) Alverno College – Case Study: Assessment in this college situated in Milwaukee is more a part of learning than actual evaluation. The term that experts associated with such assessment is horizontal integration. In this the assessment is strongly coupled with the other facets of education namely curricula design and teaching. Assessment in this case forms an integral part of daily classroom activity. The implication for us is that maybe it is not a good idea to look at assessment in isolation. This is somewhat in line with the Total Quality Management (TQM) business philosophy often used in corporate settings.

3.) Total Quality Management: A few researchers have tried to apply the TQM concepts widely used in the corporate world to schools. A specific case is the application of EFQM (European Foundation for Quality Management) framework to Pen y Dre school in US. This implied that quality was explicitly made a part of schools vision, its rules and regulations and in the teachers and staffs code of conduct [5,6].

4.) Educational Quality: Some researchers have concentrated on the very initial stuff like the definition of quality itself as it applies to education. Education is considered to meet the quality standards when it successfully provides the student with the skills that enable him/her to lead a better life both in monetary and values terms.

There are various such examples which can be quoted but what we have tried to establish through the few above examples is simply that:

Assessing quality of education through a single annual exam is probably not the best way of doing it. Quality assessment should be an overarching process which takes into account the different aspects of learning and classroom experience of a student. Hence assessment should be a continuous process much more closely tied with the overall learning process [1.2].

IV. Recommendations

Our recommendations can be broadly classified into three categories, each pertaining to an important phase of the assessment program namely creation of the assessment program its implementation and finally the generation of outputs or reports.

1.) Assessment development phase: As per our earlier discussion we feel that radical changes are required in this phase. It is important to create alternative means for integrating assessment with other learning activities. Further the activities that are used for assessment, the questions asked etc. each of these might significantly affect our outcomes, hence it is important that expert educationists' services are employed for the development of better means of evaluation. Best practices like the ones described in the previous section can be yet another way of addressing this issue.

2.) Assessment implementation phase: First and foremost it is important to segregate teachers from the assessment process. There are multiple ways in which this can be achieved. One alternative could be to outsource the on the ground assessment to a private player. This would ensure that the assessment would be free from dependency on teachers as well as the limitations of a government body like bureaucratic delays etc. But given the budget constraints of the government it may be required to develop means in which a Public Private Partnership can be set up in a mutually beneficial manner like cross subsidizing of one deal through another deal. For example, a private player would not only get access to the various Government departments and hence would become a preferred partner for various highly profit making school upgrade initiatives being taken by the Government. At the same time he would also get an understanding of the grassroots level problems and hence would be able to propose much better solutions again ensuring highly profit making contracts for remedial solutions.

Another alternative is to leverage the large graduate student bodies like NSS and NCC students for the assessment as well as the evaluation of papers. This again would ensure an unbiased assessment process while maintaining cost effectiveness. This alternative would require a new set of students to be trained every year (5-10 days of training/year). One might argue on two issues that might arise, one is the extra cost of training and the other is the credibility of these students. The first problem can be countered easily as conducting a test does not require much training and since evaluation is key-based, these extra costs would be marginal. This makes the system more reliable because teachers would not be involved in the process. The second issue is also countered in the way that the students would change every year hence no previous year biases would affect the results. Further if an OMR based assessment system is put into place then it would automatically take care of the evaluation part as the OMR sheets can be read directly by a reader attached to the computer which in turn would evaluate the responses. This would also take care of any human errors which may occur due to manual evaluation.

The third alternative is to leverage technology. We propose the use of a handheld device for the assessment process. This device which looks similar to the Leapster handheld gaming device popular in US would contain a large pool of questions and each student would be presented with a unique question paper randomly generated every time. The device would record the student's responses and communicate the same to the central server. Further this device would also contain the biometric information for the identification of students. This device would mitigate the possibilities of human error as well as malpractices while at the same time prove cost efficient in the long run.

Another important concern is the announced nature of the assessment which allows the teachers to coach their students for the assessment. The tests dates should be made unpredictable in order to ensure the students are not coached and the results represent the true picture.

3.) Results and reports phase: It is of utmost importance that the analysis conducted and reports generated by KSQAO are in-line with the expectations of the target audience. Presently KSQAO publishes an exhaustive annual report which might be too much of an information for some (like the policy makers and planners) and too less for some (like researchers). Hence KSQAO should adopt a multi level reporting format by leveraging upon organizations like National University of Educational Planning and Administration (NUEPA), which are pioneers in information management methodologies. Subsequently the analysis would also differ in order to feed relevant data to each kind of report. Further I feel that reports should be more proactive in the sense that instead of providing mere facts and figures they should provide recommendations as well a preliminary analysis indicating expecting outcomes of these recommendations.

V. CONCLUSION

KSQAO is a unique initiative started by the Karnataka State and the success it has achieved in the past 3 years is highly commendable. But there still remain some areas of concern which are stopping KSQAO from meeting its objectives. The various alternatives that this paper presented, like PPP based outsourcing model, handheld based technology solution and using NCC and NSS students for the assessment etc. are aimed at reducing the malpractices in a cost effective manner and hence ensure that KSQAO is able to produce results closer to reality.

A positive step being introduced in response to providing quality learning is KSQE (Quality Education). In the long run it hopes to see development beyond work towards an integrated learning and assessment system by adopting participatory processes involving different officers at all levels - members from teaching community, school department and monitoring committee members and members from civil society organizations. One needs to only put consistent effort towards achieving meaningful school reform to make schools work so that they churn out future citizens of our nation.

ACKNOWLEDGMENT

We would like to thank Professor S. Nayana Tara under whose guidance we undertook this project at IIM Bangalore, the officials of KSQAO and the Karnataka State Education Board who were extremely helpful in providing us all the relevant information and material and finally IIM Bangalore for providing us with the motivation and support for undertaking this project.

REFERENCES

[1] Karen Hinett, Peter Knight 1996 - Quality and assessment, Quality Assurance in Education, Volume 4, Number 3, 3-10
[2] Mantz Yorke, John Moores, 2001 – Subject benchmarking and the assessment of student learning.
[3] UNESCO, “Data Quality Assessment Framework”
[4] KSQAO {http://www.kseeb.org/Ksqao/Pages/General/Concern.htm}
[5] Viggiani, Thornett – “Quality in education – Creating a learning Scoiety”
[6] Garbutt – “Transfer of TQM from industry to education”

APPENDIX

SAMPLE IT ARCHITECTURE

PC-Based Modeling and Simulation of Large Marine Propulsion Plant

Sun Jianbo[1] Guo Chen[2]
1 Marine Engineering College, Dalian Maritime University, Dalian, Liaoning, China, 116026
E-mail: vrsjb@dlmu.edu.cn
2 Information Science and Technology College, Dalian Maritime University, Dalian, Liaoning,China,116026
E-mail: guoc@dlmu.edu.cn

***Abstract*-At present, most of marine engine room simulators (MERS) in the world are based on networking technology. But such MERS needs much money and large space to install and use. So it can not meet the demand of the training individually in spare time for the crew. So a simulation system of a PC-Based marine propulsion plant has been developed for training. In this paper, a transient simulation model for large marine propulsion system was proposed. At the same time, the marine propulsion plant control, fuel system, cooling water and etc are also developed in the simulation system. The model is a modularized structure and the modules describing system's components can be added to it easily. And related subsystems of marine propulsion plant are also established in this simulation system. The simulation experiments demonstrate the steady and dynamic results matching properly the voyage trial ones. The model can be used for marine propulsion plant performance analyzing and prediction.**

I. INTRODUCTION

For marine engine room simulator(MERS) has been listed in the Convention of STCW 78/95 as one of the essential conditions of crew training by International Maritime Organization (IMO), MERS has been a necessary equipment for modern navigation education and crew training.

The research of MERS was initiated in 1960s. In recent thirty years, MERS was developed very fast and it has become a mature technology. At present most of MERS in the world are based on networking technology, which is composed of many computers that realize different functions of different systems. But the simulation system based on networking technology not only need vast resources such as large space and much money, but also is inconvenient to install and use. So it is unfit for crew to understand the related knowledge of the diesel engine and to make a practice individually in spare time.

The simulation system of a PC-based diesel engine could be used as a necessary complement of MERS based on networking technology. And it is suitable for the crew to be familiar with the operation of system, to comprehend all the systems of the diesel engine, to complete the related knowledge training. Its has advantages of lower investment, less occupied space, lower requirement of system configuration, more clear system structure, easier to install and use. The simulation system, which can meet the need of practice individually in spare time, is one of the most important parts of marine engine operation training.

II. DIESEL ENGINE PROPULSION SYSTEM MODEL

A. *Marine Propulsion Plant Model*

Fig.1. shows the block diagram of the marine propulsion system mathematical model.

The dynamic characteristics of diesel engine speed are in the left of Fig.1. And the diesel engine speed can be calculated by the following differential equation.

$$\frac{\pi}{30} I \frac{dn}{dt} = M_{shaft} - M_{prop} \tag{1}$$

Where I is the equivalent inertia moment of power plant (considering the effect of incidental water), n is the speed of diesel engine, M_{shaft} is the output torque, M_{prop} is the loading torque of the propeller.

The dynamic characteristics of hull are in the right of Fig.1. Ship hull motion equation should be 3-d motion equation. To simplify model, only consider the motion of x direction and rudder angle is zero, so the equation of ship motion is as follows:

$$(m_{ship} + m_w)\frac{dV_s}{dt} = T_{prop} - R_{ship} \tag{2}$$

Where m_{ship} and m_w are the ship mass and additional water mass; T_{prop} is the effective thrust of the propeller; R_{ship} is the hull resistance.

A regression algorithm is used to calculate the propeller thrust and torque in order to avoid complex calculation of the propeller's characteristics.

Define β is

$$\beta = \arctan\left(\frac{V_a}{k_1 \pi n D}\right) \tag{3}$$

Where V_a is the propeller's advance; D is the diameter of the propeller; k_1 is coefficient.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_70, © Springer Science+Business Media B.V. 2010

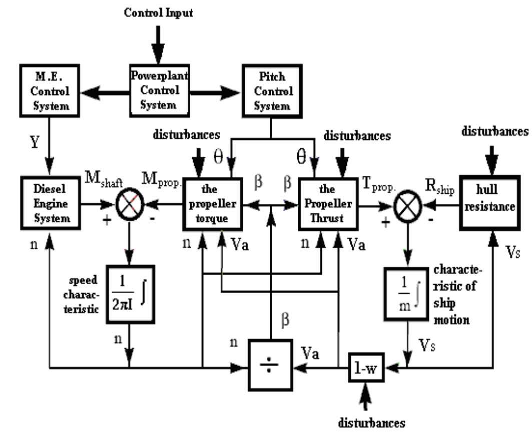


Fig.1. Block diagram of the marine propulsion system mathematical model

So the propeller thrust coefficient $C_T^{\,*}$ and torque coefficient $C_Q^{\,*}$ in open-water could be calculated by using equations as follows:

$$C_T^{\,*} = \frac{1}{100}\sum_{k=1}^{30}\{A(k)\cos(k\beta) + B(k)\sin(k\beta)\} \tag{4}$$

$$C_Q^{\,*} = -\frac{1}{1000}\sum_{k=1}^{30}\{A(k)\cos(k\beta) + B(k)\sin(k\beta)\} \tag{5}$$

And the calculation equations of the propeller's effective thrust T_{prop} and loading torque M_{prop} are given as following:

$$T_{prop} = \frac{\pi K_1}{8}(1-t)C_T^{\,*}\rho\{V_a^{\,2} + (k_2\pi n D)^2\}D^2 \tag{6}$$

$$M_{prop} = \frac{\pi K_2}{8}C_Q^{\,*}\rho\{V_a^{\,2} + (k_2\pi n D)^2\}D^3 \tag{7}$$

Where ρ is seawater density; t is thrust deduction fraction coefficient; K_1 and K_2 are correction coefficients which are influenced on by navigation condition; k_2 is coefficient;

While the propeller's effective thrust, loading torque and hull resistance are calculated, the influence of various disturbances (e.g. wind, wave, shallow waterway, narrow waterway and etc.) also need to be considered; k_1 in (3) and k_2, K_1 and K_2 in (6), (7) are the coefficients that are influenced on by navigation condition. Thus, the dynamic characteristic of propulsion power plant under rough sea condition could be realized to meet the need of simulation system.

B. Main diesel engine model

Fig.2. shows block diagram of a low-speed two-stroke diesel engine quasi-steady model [1-3]. Based on the modeling method of quasi-steady, diesel engine could be decomposed to turbocharger (including air compressor and exhaust turbine), air cooler, scavenge box, cylinder, exhaust pipe and etc. And the models of those modules are established [4].

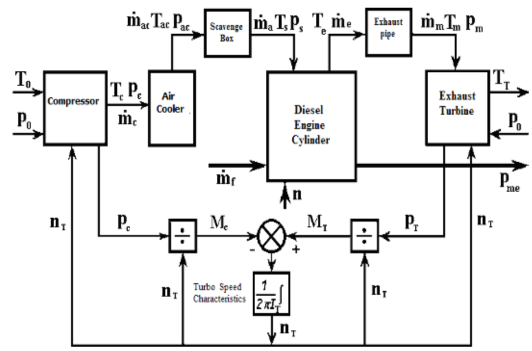


Fig. 2 Block diagram of the diesel engine model

C. Marine Propulsion Control System Model

Auto Chief 4(AC4) is diesel engine fixed-pitch propeller propulsion control system, which has features of strong currency, stability, reliability, convenient operation, developed by Company Norcontrol Automation AS in Norway. And it is widely used on merchant ship in the world.

DGS 8800eType digital governor system is the key of AC4 propulsion control system. The model of digital governor system includes a regulation of main engine speed by PID regulating algorithms and a load controller (scavenges pressure limiter, torque limiter and manual limiter). And Fig.3 is simulation model diagram of DGS 8800e digital governor.

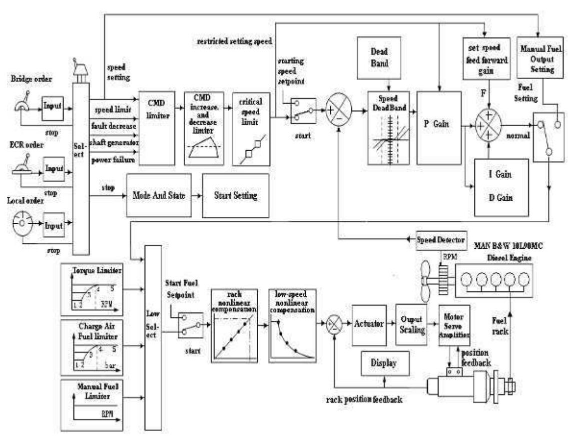


Fig. 3 Simulation Model Diagram of DGS8800e Digital Governor

From the diagram, the model of DGS 8800eType digital governor mainly includes the module of setting speed restriction, the module of acceleration and deceleration restriction, the module of critical speed restriction, the module of starting setting, the module of PID controller, the module of manual fuel setting, the module of load restriction, the module of compensation, the regulator servomechanism and etc.

D. *Subsystem Model*

A variety of subsystem models have been updated /extended or numerically configured. The models included:

a) Fuel oil system model
b) Lubrication Oil system model
c) Cooling water system model
d) Compression air system model
e) Boiler and Steam system model
f) Bilge Oil and bilge water system model
g) Purifier system model

III. REALIZATION OF SIMULATION SYSTEM

Based on the modularized modeling method, the whole of simulation system is decomposed into different function modules. And the models of the main diesel engine and marine propulsion system , main engine control system, compressed air system, fuel and lubrication oil purification system , fuel oil system, lubrication oil system, cooling water system, patrolling and alarm system, steam system and etc. are established; And the corresponding simulation interfaces are also realized. Fig.4 shows block diagram of simulation system.

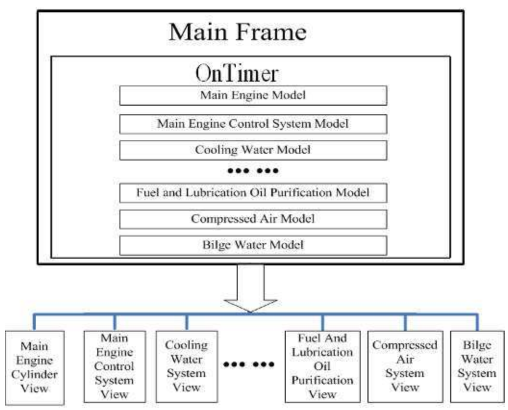


Fig.4. Block diagram of the simulation system

While designing the program, in order to reduce overhead of memory, to lighten the tasks of CPU and to realize real time simulation, the simulation interfaces all use the frame structure with single document and single view. Every operation interface has its own view, which is switched through the function of switching view in main frame. When a view is switched, the view is not just hidden (if being hidden, the view will occupy the resource of memory), it is destroyed. So there is only one view in memory, thus is more effective in saving memory, improving processing speed and reducing the requirement of the configuration of a PC. So it meets the requirement of the designing based on a PC.

The interfaces of DGS8800e digital governor operation panel and engine control room are shown in Fig.5 and Fig.6.

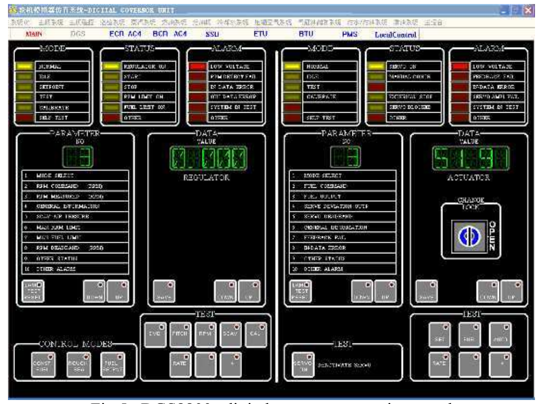


Fig.5. DGS8800e digital governor operation panel

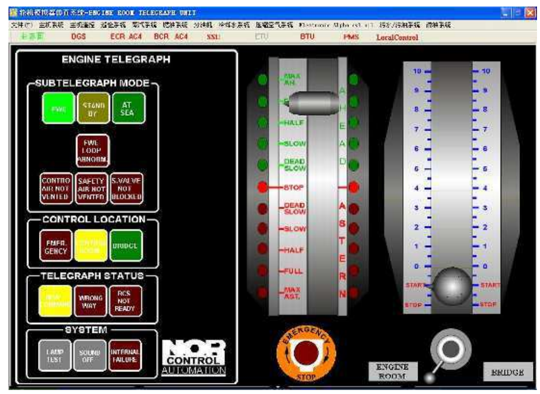


Fig.6 Engine control room unit

Fig.7 shows dynamic simulation interface of the Pneumatic maneuvering system of main diesel engine. Operator can fulfill stand-by, opening or closing the valves as well as watching dynamic processes of the starting, reversing and stopping the main diesel engine.

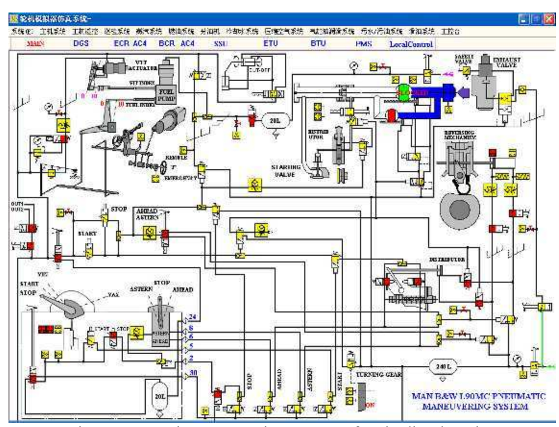


Fig.7. Pneumatic maneuvering system of main diesel engine

IV. *RESULTS OF SIMULATION*

MAN B&W 10L90MC diesel engine, which is the main propulsion power plant of the fifth 5446ETU large container ship "Cosco Rotterdan", is chosen as the object for the simulation. Principal specifications of the diesel engine are given in table I. and the propeller particular is presented in table Π.

TABLE I
PARTICULARS OF DIESEL ENGINE

Type	MAN B&W 10L90MC
Number of cylinders	10
Cylinder diameter	0.9 m
Stroke	2.916 m
Connecting rod length	3.510 m
Compression ratio	17.8
MCR speed	82 r/min
MCR power	43100 kW
Max. combustion pressure	13.5-14 MPa
BMEP	1.74 MPa
Fire order (for Ahead)	1-8-7-3-5-9-4-2-10-6
Turbochargers	3 NA 70/T9018

TABLE Π
PARAMETER OF THE PROPELLER

Number of blades	5
Diameter	9.200 m
Pitch(at 0.7R)	9.657 m
Exp. Area	44.4 m^2

A regression algorithm is used to calculate the propeller thrust and torque in order to avoid complex calculation of the propeller's characteristics. Fig 8 shows the results of the torque coefficient C_Q and thrust coefficient C_T.

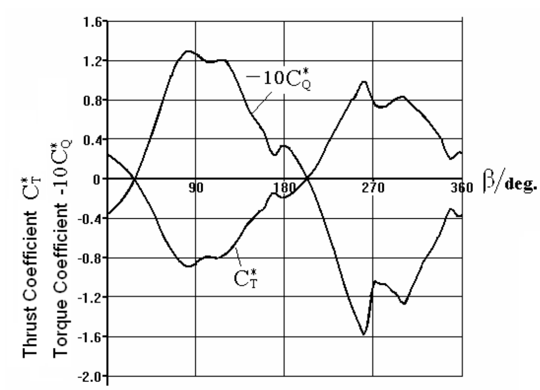


Fig.8 Relation between torque coefficient C_Q, thrust coefficient C_T and β

Steady-state simulation runs at 100.8%, 90.8%, 75.4%, 70.6% and 50.5% loads were performed. The results derived from the steady-state simulation were that the engine steady-state performance is predicted with enough accuracy to allow the simulation of transient phenomena with confidence (fig.9).

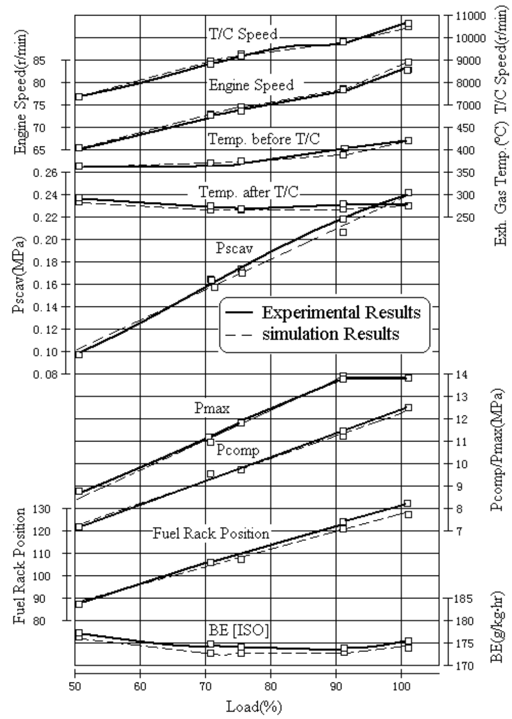


Fig. 9 Comparison of experimental results to simualtion: steady-state several loads

As an example, the case the engine ahead running initially at 84 rpm with ship speed 27 knots suddenly reverse to crash astern will was chosen and calculated. The results of the crash astern prediction are given in Fig.10. These include the ship speed, the ship sailing distance and the engine speed. The simulation results show the simulation results are well accord with the trail voyage data.

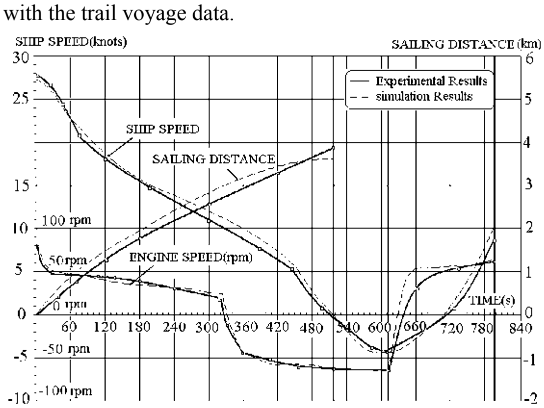


Fig.10 Results of the transient simulation at crash astern operation condition

Transient simulation runs were performed. In the first run, changes in the engine speed setpoint from 84 to 40 rpm and from 40 to 84 rpm were applied at the 10th and 60th second of the run, respectively. The results of transient simulation runs are given in Fig.11.

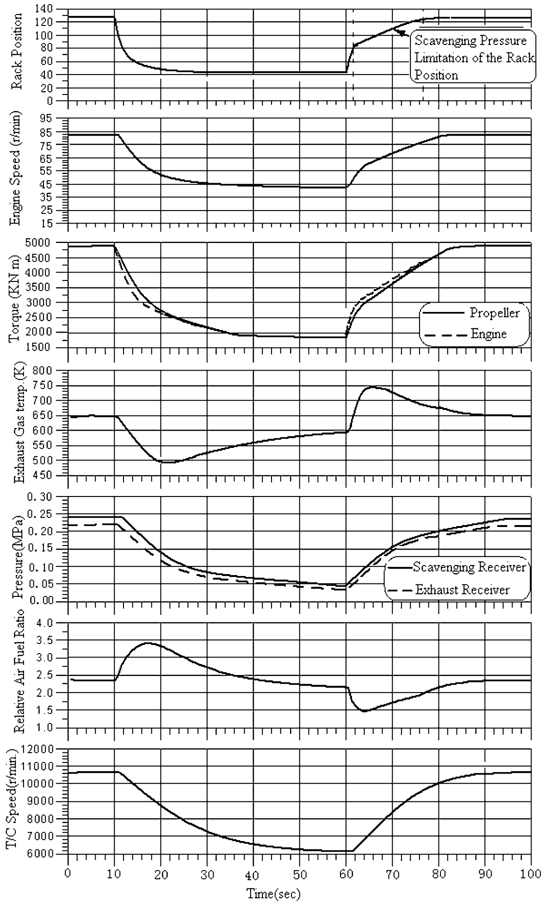


Fig.11 The transient simulation-change in engine speed setpoint from 84 to 40 rpm and from 40 to 84 rpm

V. CONCLUSIONS

1) A real-time simulation model of a large low-speed two-stroke marine diesel engine and the propeller is proposed. And the results of simulation under different working conditions are given in the paper. The simulation experiment shows that the model reflects the dynamic and steady-state characteristics accurately.

2)A simulation software of marine propulsion plant based on a personal computer is developed, which includes main engine system, marine propulsion control system, compression air system, Purifier system, fuel oil system, lubrication oil system, cooling water system, patrolling and alarm system, steam system and etc.

The simulation system of a PC-based marine propulsion plant could be used as the reinforcement of MERS based on networking technology. It has advantages of lower investment, less occupied space, lower requirement of system configuration, easier to install and use. The simulation system can meet the need of practice individually in spare time, and is suitable for the crew to complete the related knowledge training.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No. 60774046). The authors would like to thank the crew of the vessel “Cosco Rotterdan” giving experimental results measured on the MAN B&W 10L90MC diesel engine available for model validation purposes.

REFERENCES

[1] Wan-Chung LAN, Takeshi KATAGI. Quasi Steady State Simulation of Diesel Engine Transient Performance and Design of Machatronic Governor. Bulletin of the M.E.S.J.1996, vol. 24:1-13.

[2] J.Arregle, J.J. Lopez. Development of a zero-dimensional diesel combustion model. Part 1: Analysis of the quasi-steady diffusion combustion phase. Applied Thermal Engineering, 2003, vol. 23, pp.1301-1317.

[3] Blanke M, Andersen J.S. On Modeling Large Two-Stroke Diesel Engine: New Results from Identification. Budapest: IFAC 9th Triennial World Congress, 1984, pp.2015-2020.

[4] Sun Jianbo. Quasi Steady State Simulation of Large-Bore Low-Speed Two-Stroke Diesel Engine Propulsion Plant System. Tianjin: Transaction of CISCE , 2006, vol. 24, pp.351-357.

Evolutionary Techniques for Deterministic Chaos Control

Roman Senkerik, Ivan Zelinka, Zuzana Oplatkova
Faculty of Applied Informatics, Tomas Bata University in Zlin
Nad Stranemi 4511, Zlin 762 72, Czech Republic
{senkerik, zelinka,oplatkova}@fai.utb.cz

Abstract— **This work deals with optimization of the control of chaos by means of evolutionary algorithms. The main aim of this paper is to show that powerful optimizing tools like evolutionary algorithms can be in reality used for the optimization of deterministic chaos control. This work is aimed on explanation of how to use evolutionary algorithms (EAs) and how to properly define the cost function (CF). It is also focused on selection of control method and the explanation of all possible problems with optimization which comes together in such a difficult task, which is chaos control. Five new approaches for constructing the cost functions leading to satisfactory results are presented here. These cost functions secure fast and precise stabilization of chaotic system. As a model of deterministic chaotic system, the one dimensional logistic equation was used. The evolutionary algorithm Self-Organizing Migrating Algorithm (SOMA) was used in four versions. At the end, the brief overview of most important results is also presented.**

I. Introduction

Currently, evolutionary algorithms (EA) are known as powerful tools for almost any difficult and complex optimization problem. But the quality of obtained through the optimizations mostly depends on proper design of the used cost function, especially when the EAs are used for optimization of chaos control. It is well known that deterministic chaos in general and also any technique to control of chaos are sensitive to parameter setting, initial conditions and in the case of optimization they are also extremely sensitive to the construction of used cost function. This research utilized the Pyragas's delayed feedback control technique [1], [2]. Unlike the original OGY control method [3] it can be simply considered as a targeting and stabilizing algorithm together in one package [4]. Another big advantage of Pyragas method is the amount of accessible control parameters. This is very advantageous for successful use of optimization of parameters setting by means of EA, leading to improvement of system behavior and better and faster stabilization to the desired periodic orbits. Some research in this field has been recently done using the evolutionary algorithms for optimization of local control of chaos [5], [6]. Also the question of targeting (faster stabilization) with application to chaos control has attracted the researchers. The several first approaches for targeting have used special versions of OGY control scheme [7], [8] or collecting of information about trajectories, which fall close to desired state [9]. Later, lots of methods were based on adaptive approach [10], center manifold targeting [11] or neural networks [12], [13].

The control law in this work is based on Pyragas method: Extended delay feedback control – ETDAS [14]. This research is a continuation of previous experiments with application of EA to chaos control [15 - 17].

This work was supported by the grant NO. MSM 7088352101 of the Ministry of Education of the Czech Republic and by grants of Grant Agency of Czech Republic GACR 102/06/1132.

II. Problem Design

The chosen example of chaotic systems was one-dimensional Logistic equation in form (1).

$$x_{n+1} = rx_n\left(1 - x_n\right) \tag{1}$$

The logistic equation (logistic map) is a one-dimensional discrete-time example of how complex chaotic behavior can arise from very simple non-linear dynamical equation. This chaotic system was introduced and popularized by the biologist Robert May [18]. It was originally introduced as a demographic model as a typical predator – prey relationship. The chaotic behavior can be observed by varying the parameter r. At r = 3.57 is the beginning of chaos, at the end of the period-doubling behavior. At r 3.57 the system exhibit chaotic behavior

The example of this behavior can be clearly seen from bifurcation diagram – Fig. 1.

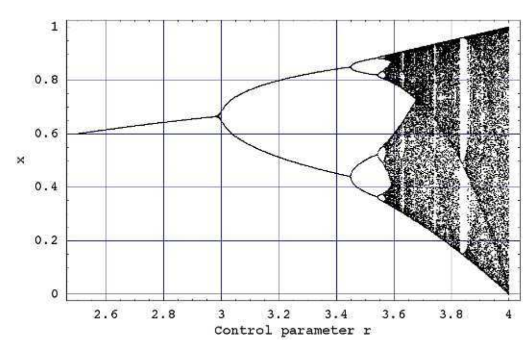


Fig. 1. Bifurcation diagram of Logistic equation

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_71, © Springer Science+Business Media B.V. 2010

This work is focused on explanation of application of EA for estimation of three accessible control parameters for EDTAS method to stabilize desired Unstable Periodic Orbits (UPO) and comparison of results for presented cost functions. Desired UPOs are following: p-1 (a fixed point) case and p-2 (higher periodic orbit – oscillations between two values). All simulations were 50 times repeated for each EA version to show and check robustness of used method. The control method – ETDAS in the discrete form suitable for one dimensional logistic equation has the form (2).

$$
\begin{aligned}
x_{n+1} &= r x_n (1 - x_n) + F_n \\
F_n &= K\left[(1 - R) S_{n-m} - x_n\right] \\
S_n &= x_n + R S_{n-m}
\end{aligned}
\tag{2}
$$

Where K and R are adjustable constants, F is the perturbation, S is given by a delay equation utilizing previous states of the system and m is the period of m-periodic orbit to be stabilized. The perturbation F_n in equations (2) may have arbitrarily large value, which can cause diverging of the system outside the interval $\{0, 1.0\}$. Therefore, F_n should have a value between $-F_{\max}$, $F_{\max}$ and EA should find an appropriate value of this limitation to avoid diverging of system.

III. THE COST FUNCTION

Several unique designs of cost functions (CF) were developed and tested for stabilization of p-1 orbit (fixed point), and p-2 orbit. The CF has been calculated in general from the distance between desired state and actual system output. The minimal value of this cost function revealing the best solution is zero. The aim of all the simulations was to find the best solution that returns the cost function value as close as possible to zero.

A. Basic CF

This proposal of the basic cost function is in general based on the simplest CF, which could be used only for the stabilization of p-1 orbit. The idea was to minimize the area created by the difference between the required state and the real system output on the whole simulation interval – $\square_i$.

But another cost function (CF) had to be used for stabilizing of higher periodic orbit. It was synthesized from the simple CF and other terms were added. In this case, it is not possible to use the simple rule of minimizing the area created by the difference between the required and actual state on the whole simulation interval – $\square_i$, due to the many serious reasons, for example: degrading of the possible best solution by phase shift of periodic orbit.

This CF is in general based on searching for desired stabilized periodic orbit and thereafter calculation of the difference between desired and found actual periodic orbit on the short time interval - $\square_s$ (approx. 20 - 50 iterations) from the point, where the first min. value of difference between desired and actual system output is found. Such a design of CF should secure the successful stabilization of higher periodic orbit anywise phase shifted.

Furthermore, because of CF values being very close to zero, this CF also allows using of decision rule avoiding very time demanding simulations. This rule stops EA immediately, when the first individual with good parameter structure is reached, thus the value of CF is lower then the acceptable (CF_{acc}) one. Typically CF_{acc} = 0.001 at time interval $\square_s$ = 20 iterations, thus difference between desired and actual output has value 0.0005 per iteration – i.e. successful stabilization for used control technique. This CF was also used for p-1 orbit. The CF_{Basic} has the form (3).

$$
CF_{Basic} = penalization_1 + \sum_{t=\tau 1}^{\tau 2} \left| TS_t - AS_t \right| \tag{3}
$$

where: TS - target state, AS - actual state

τ_1 - the first min. value of difference between TS and AS

τ_2 – the end of optimalizing interval ($\tau_1 + \tau_s$)

penalization$_1$= 0 if τ_i - $\tau_2 \geq \tau_s$;

penalization$_1$= 10*(τ_i - τ_2) if τ_i - $\tau_2 < \tau_s$ (i.e. late stabilization)

B. Targeting CF Simple

In this case study the simplest CF proposal outlined above was used. It is based on minimizing the area created by the difference between the required state (stabilized fixed point) and the real system output on the whole simulation interval – $\square_i$, thus this proposal of CF should secure fast targeting into the close neighborhood of p-1 orbit and its stabilization. The CF_{Simple} is given by (4).

$$
CF_{simple} = \sum_{t=0}^{\tau_i} \left| TS_t - AS_t \right| \tag{4}
$$

where: TS - target state, AS - actual state

This CF was used only for optimization in the task of improvement of stabilization of p-1 orbit. Following proposal of CF NA was used for both p-1 and p-2 orbit.

C. Targeting CF NA

It was necessary to modify the definition of CF in order to decrease the average number of iteration required for successful stabilization and avoidance of any associated problem. The CF_{Simple} is not suitable for adding any term of penalization for slowly stabilizing solutions, thus the CF_{Basic}

was modified to use for all required UPOs. The CF value is multiplied by the number of iterations (*NI*) of the first found minimal value of difference between desired and actual system output (i.e. the beginning of fully stabilized UPO). To avoid problems associated with CF returning value 0 and to put the penalization to similar level as the non-penalized CF value, the small constant (*SC*) is added to CF value before penalization (multiplying by *NI*). The modified CF_{NA} has the form (5).

$$CF_{NA} = NI\left(SC + penalization1 + \sum_{t=\tau 1}^{\tau 2} \left| TS_t - AS_t \right| \right) \qquad (5)$$

where: $SC = 10^{-16}$ for p-1 orbit,
$SC = 10^{-8}$ for p-2 orbit

D. Targeting CF Targ1

The next proposal of CF design is based on the previous one with small change, which should avoid any problems with defining the value of small constant *SC* in advance (especially for stabilization of higher periodic orbit). The *SC* value (7) here is computed with the aid of power of non-penalized basic part of CF.

In general, there exists two possible ways for applying the multiplication by number of iterations required for stabilization (*NI*). The first version of final design of targeting CF (CF_{TARG1}) has the form (6). Here the sum of basic part of CF and automatically computed *SC* is multiplied by *NI*. Consequently, the EA should find the solutions securing the fast targeting into desired behavior of system.

$$CF_{TARG1} = NI\left(SC + penalization1 + \sum_{t=\tau 1}^{\tau 2} \left| TS_t - AS_t \right| \right) \qquad (6)$$

where:

$$SC = 10^{\text{Expcf}}, \; Expcf = \log_{10}\left(\sum_{t=\tau 1}^{\tau 2} \left| TS_t - AS_t \right| + 10^{-15} \right) \qquad (7)$$

E. Targeting CF Targ2

In the second version of targeting CF (CF_{TARG2}), there is only slight change in comparison with the previous proposal. Here the number of steps for stabilization (*NI*) multiplies only the small constant (*SC*) which is counted in the same way as in the previous case (7) This version of targeting CF (CF_{TARG2}) has the form (8)

$$CF_{TARG2} = (NI \cdot SC) + penalization1 + \sum_{t=\tau 1}^{\tau 2} \left| TS_t - AS_t \right| \qquad (8)$$

F. CF Overview

The results given by these five CFs (3) – (8) are outlined later. The difference between them can be clearly seen in Figs. 2 - 4, which show the dependence of CF values on the adjustable parameters K and F_{max}. From these figures, it is obvious that together with growing complexity of used CF, grows also the nonlinearity and unpredictability of CF surface. Thus this is the answer for the question why the EA were used.

The illustrative examples related to CF_{TARG1} and CF_{TARG2} are not presented here due to the close graphical similarity with CF_{NA} (Fig. 4)

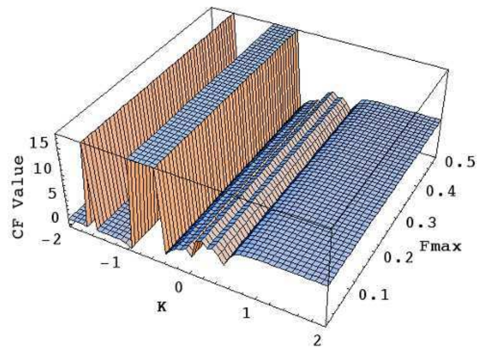


Fig. 2. Dependence of CF value on parameters K and F_{max} for p-1 orbit, $x_{initial} = 0.8$, $R = 0$, CF Basic

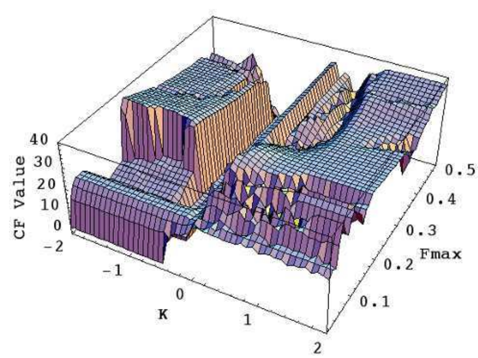


Fig. 3. Dependence of CF value on parameters K and F_{max} for p-1 orbit, $x_{initial} = 0.8$, $R = 0.0180$, CF Simple

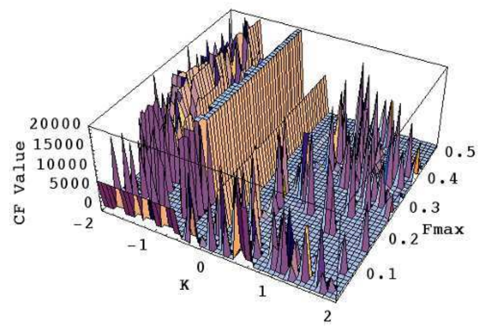


Fig. 4. Dependence of CF value on parameters K and $F_{\max}$ for p-1 orbit, $x_{initial} = 0.8$, $R = 0.4977$, CF NA

IV. Optimizing Algorithm

For the experiments described here, stochastic optimization algorithm SOMA [19] has been used. It was chosen because it has been proven that this algorithm has the ability to converge towards the global optimum. SOMA works with groups of individuals (population) whose behavior can be described as a competitive – cooperative strategy. The construction of a new population of individuals is not based on evolutionary principles (two parents produce offspring) but on the behavior of social group, e.g. a herd of animals looking for food. This algorithm can be classified as an algorithm of a social environment. To the same group of algorithms, particle swarm algorithm can also be put in, sometimes called swarm intelligence. In the case of SOMA, no velocity vector works as in particle swarm algorithm, only the position of individuals in the search space is changed during one generation, here called 'Migration loop'.

The rules are as follows: In every migration loop, the best individual is chosen, i.e. individual with the minimum cost value, which is called "Leader". An active individual from the population moves in the direction towards Leader in the search space. At the end of the movement, the position of the individual with minimum cost value is chosen. If the cost value of the new position is better than the cost value of an individual from the old population, the new individual appears in new population. Otherwise the old one remains for the next migration loop.

V. Experimental Results

Four versions of SOMA were used for all simulations (See Table 1). Parameters for the optimizing algorithm were set up in such a way in order to reach the same value of maximal CF evaluations for all used versions (See Table 2). Each version of SOMA has been applied 50 times in order to find the actual optimum.

The primary aim here is not to show which version is better or worse but to show that the EA can in reality be used for deterministic chaos control when the cost functions are properly defined and to show the brief overview of results for the presented five CF designs. Here is the list of desired UPOs:.

Logistic Equation with $r = 3.8$:
p-1 (fixed point): $x_F = 0.73842$
p-2 orbit: $x_1 = 0.3737$, $x_2 = 0.8894$

The optimization interval for p-1 orbit was □i = 100 iterations, for higher periodic orbits it was mostly □i = 150 iterations.

The ranges of all estimated parameters were these:

$$-2 \leq K \leq 2\ ,\ 0 \leq F_{\max} \leq 0.5 \text{ and } 0 \leq R \leq 0.5$$

TABLE I
USED VERSIONS OF SOMA

Index	Algorithm / Version
1	SOMA AllToOne
2	SOMA AllToRandom
3	SOMA AllToAll
4	SOMA AllToAllAdaptive

TABLE II
PARAMETER SETTINGS FOR SOMA

Parameter	ATO / ATR	ATA / ATAA
PathLength	3	3
Step	0.33	0.33
PRT	0.1	0.1
PopSize	25	10
Migrations	25	7
Max. CF Evaluations (CFE)	5400	5670

For the comparison of average number of iterations required for successful stabilization (Istab value) (all 50 repeated simulations), please refer to Table 3. Here it can be clearly seen the gradual decreasing of average Istab value together with development and testing of complex targeting cost functions.

TABLE III
RESULTS FOR P-1 AND P-2 ORBIT

UPO	p-1	p-2
CF Basic	77	124
CF Simple	65	-
CF NA	49	114
CF Targ1	49	113
CF Targ1	39	108

This section also contains graphics representation of the results given by optimizations by means of CF Targ2 (8) developed on the basis of successful CF NA design.

The figures shows the simulation of the best individual solutions given by SOMA under identical initial conditions used in optimization and the complex simulation for the uniformly distributed initial conditions in the range $0 < x_{initial} < 1$, 100 samples were used in this kind of simulation.

The presented Figs. 5 – 8 lend weight to the argument, that this CF design is a serious consideration in the robust stabilization of chaotic systems and seem to be the best choice for the task of finding of "universal and robust solution". The disadvantage of this proposal is relatively big computational-time demands.

Another tiny disadvantage of this chaos control approach is the occurrence of the phenomenon, that in case of higher periodic orbits some best individual solutions are not suitable for complex simulation with wide range of initial conditions, whereas remaining best solutions give excellent results.

As can be seen from the depicted simulation results, the CF Targ2 design gives excellent performance from the point of view of quickness and quality of stabilization. This CF secures very fast targeting to the desired state.

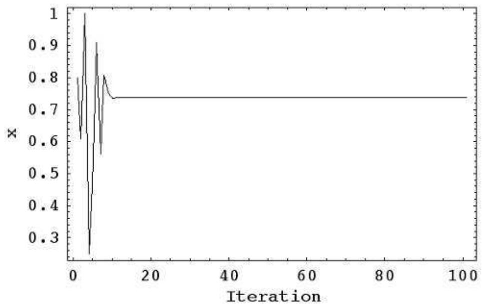


Fig. 5. Best ind. solution: p-1 orbit, CF Targ2, SOMA ATO

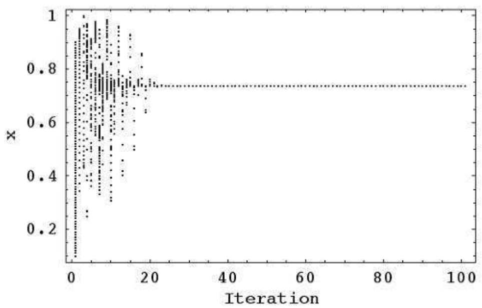


Fig. 6. Best solution: p-1 orbit, CF Targ2, SOMA ATO complex simulation with $0 < x_{initial} < 1$.

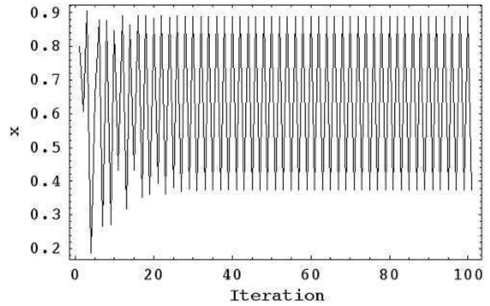


Fig. 7. Best ind. solution: p-2 orbit, CF Targ2, SOMA ATO

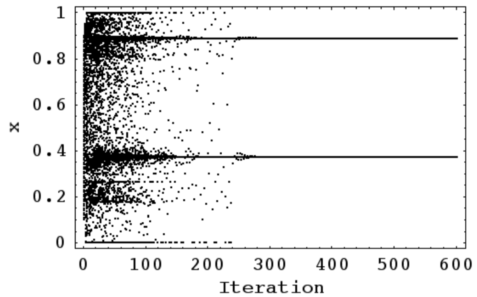


Fig. 8. Best solution: p-2 orbit, CF Targ2, SOMA ATAA, complex simulation with $0 < x_{initial} < 1$.

VI. Conclusions

The optimization of chaos control described here is relatively simple and easy to implement. Based on obtained results, it may be claimed that all simulations give satisfactory results and thus EAs are capable of solving this class of difficult problems and the quality of results does not depend only on the problem being solved but also on the proper definition of the CF.

From the optimization results follows, that they are extremely sensitive to the construction of used CF and any small change in the design of CF can cause radical improvement of system behavior (as in case of CF Targ2), but of course on the other hand can cause worsening of observed parameters and behavior of chaotic system as well.

All achieved results give the following piece of knowledge.

The first proposed CF Basic gives satisfactory results and can be used wherever the good quality of stabilization is expected and the speed of stabilization and "universality of this solution" for wider range of initial conditions are not decisive. This CF does not require any special experiences and knowledge about the system.

The second CF simple represents the simplest example of targeting CF suitable only for stabilization of fixed point. In comparison with CF Basic it gives similar results.

The next proposal of CF NA represents the progressive targeting CF suitable for p-1 and p-2 orbit, which gives very good results in the task of shortening of the initial chaotic stage. The results for p-1 orbit are significantly better than in the previous two CFs, on the other hand the slightly better results for p-2 orbit were achieved at the cost of arising of problem with worse performance of EAs and obtaining of solutions with only temporary stabilization or none at all. Moreover this CF requires some knowledge about results achieved in the case of CF Basic due to proper setting of *SC* value.

The fourth CF Targ1 brings the advantage of automatically computed *SC* value, thus it is suitable for any desired UPO. The obtained results were similar as in case of CF NA.

In the last proposal of CF Targ2 there were only slight changes in CF design, but from the presented results it can be seen, how such a small change can influence the performance of a controlled system, especially when it is an extremely sensitive chaotic system. Here, another improvement from the point of view of quickness of stabilization was achieved and furthermore the performance of EAs was increased. This seems to be the best choice, when very good solution from the close neighborhood of initial conditions is expected.

Finally, it is hard task to propose a CF, which gives excellent results, especially "universal results" suitable for simulation with wide range of initial conditions. As repeatedly mentioned the chaotic systems are extremely sensitive to proper settings of control algorithm and of course they are very sensitive to even very tiny change in any parameter. This extreme sensitivity is transferred into complexity of CF surface thus it is also hard task for EAs to find good solution. It is also difficult to determine the conditions for optimizations and subsequent simulations.

As a consequence of these facts it is possible to say that all presented CFs gives good results and each one is more or less suitable depending on concrete demands for quickness or quality of stabilization, computational-time demands, order of UPO, whether it should be the solution only for limited circle of initial conditions or for wider range etc.

There is no problem for the future research in defining much more complex CF comprising as subcriteria control of stability, costs, time-optimality, controllability, or any of their arbitrary combinations. Furthermore parameter settings for EA were based on heuristic approach; therefore there is also possibility for the future research. According to all results showed here it is planned that the main activities will be focused on testing of evolutionary deterministic chaos control in continuous-time and high-order systems and finally testing of evolutionary real-time chaos control.

REFERENCES

[1] Just W., 1999, "Principles of Time Delayed Feedback Control", In: Schuster H.G., *Handbook of Chaos Control*, Wiley-Vch, ISBN 3-527-29436-8.

[2] Pyragas K., 1992, "Continuous control of chaos by self-controlling feedback", *Physics Letters A*, 170, 421-428.

[3] Ott E., C. Greboki, J.A. Yorke, "Controlling Chaos", *Phys. Rev. Lett.* vol. 64, 1990, pp. 1196-1199.

[4] Kwon O. J., "Targeting and Stabilizing Chaotic Trajectories in the Standard Map", *Physics Letters A*. vol. 258, 1999, pp. 229-236.

[5] Richter H., K. J. Reinschke, "Optimization of local control of chaos by an evolutionary algorithm", *Physica D*, vol. 144, 2000, pp. 309-334.

[6] Richter H., "An Evolutionary Algorithm for Controlling Chaos: The Use of Multi - Objective Fitness Function", *Lecture Notes in Computer Science*, vol. 2439, 2002, pp. 308-320.

[7] Bollt E. M., Kostelich E. J., „Optimal Targeting of Chaos", *Physics Letters A*. vol. 245, 1998, pp. 399-406.

[8] Kostelich E. J., Ott E., Grebogi C., Yorke J. A., „Higher-dimensional Targeting", *Physical Review E*. vol. 47, No 1, 1993, pp. 305-310.

[9] C. M. Bird, P. J. Aston," Targeting in the Presence of Noise", *Chaos, Solitons, & Fractals*, Volume 9, Issues 1-2, 1998, Pages 251-259.

[10] Ramaswamy R., Sinha S., Gupte N., „Targeting Chaos Through Adaptive Control", *Physical Review E*. vol. 57, No 3, 1998, pp. 2507-2510.

[11] Starrett J., „Time-optimal Chaos Control by Center Manifold Targeting", *Physical Review E*. vol. 66, No 4, 2002, pp. 6206-6211.

[12] S. Iplikci, Y. „Denizhan, Control of chaotic systems using targeting by extended control regions method", *Physica D*: Nonlinear Phenomena, Volume 150, Issues 3-4, 1 April 2001, Pages 163-176.

[13] S. Iplikci, Y. Denizhan, „An improved neural network based targeting method for chaotic dynamics", *Chaos, Solitons & Fractals*, Volume 17, Issues 2-3, July 2003, Pages 523-529.

[14] Pyragas K., "Control of chaos via extended delay feedback", *Physics Letters A*, vol. 206, 1995, pp. 323-330.

[15] Zelinka I., Senkerik R., Navratil E.,"Investigation on Evolutionary Optimitazion of Chaos Control", *CHAOS, SOLITONS & FRACTALS* (2007), doi:10.1016/j.chaos.2007.07.045

[16] Senkerik R., Zelinka I., Navratil E., "Optimization of feedback control of chaos by evolutionary algorithms", in *proc 1st IFAC Conference on analysis and control of chaotic systems*, Reims, France, 2006, In Press.

[17] Zelinka I., Senkerik R., Navratil E., "Investigation on Real Time Deterministic Chaos Control by Means of Evolutionary Algorithms", CHAOS'06, *In Proc. 1st IFAC Conference on Analysis and Control of Chaotic Systems*, Reims, France, 28-30 June 2006, pages 211-217.

[18] May R.M., "Stability and Complexity in Model Ecosystems", Princeton University Press, 2001, ISBN: 0-691-08861-6.

[19] Zelinka I., "SOMA – Self Organizing Migrating Algorithm", In: *New Optimization Techniques in Engineering*, (B.V. Babu, G. Onwubolu (eds)), chapter 7, 33, Springer-Verlag,

Active Vision for Human Action Sensing

M. Mackay, R. G. Fenton, and B. Benhabib
Department of Mechanical & Industrial Engineering
University of Toronto, 5 King's College Road
Toronto, ON, Canada, M5S 3G8

Abstract – **This paper presents a *real-world* implementation of a multi-camera active vision system for form/action recognition of time-varying geometry objects (specifically, the human form). The proposed method uses a novel sensing-system reconfiguration methodology to improve recognition performance. Near-optimal view-points are selected by the algorithm, along a prediction time horizon. to maximize target visibility in the presence of multiple, potentially moving obstacles. The proposed reconfiguration method has been tested using controlled experiments.**

I. Introduction and Literature Review

Vision-based human form and action recognition typically has been attempted under ideal conditions, i.e., with clear, un-occluded views of the subject [1]. Mobile (active) sensors can significantly improve recognition performance by reducing uncertainty – views from optimal viewpoints ease the requirements on the recognition algorithm [2], [3]. Herein, an agent-based sensing-system reconfiguration methodology is proposed to improve system performance in recognizing a human walking motion in the presence of obstacles and a cluttered environment.

A. Sensing-System Reconfiguration

One can define sensing-system reconfiguration as the use of a formal method to select the number, types, locations, and internal parameters of sensors employed in the surveillance of a target [4]. Most papers classify reconfiguration methods as either *generate-and-test* or *synthesis* [1]. The former methods discretize the sensor-pose domain, for example, by using a tessellated sphere or geodesic dome [5]. Synthesis methods determine sensor parameters by finding an analytical solution to task constraints in a high-dimension space ([6]).

Dynamic environments, with moving targets, obstacles, and sensors have also been considered [7], [11]. Many are simply extensions of static-environment methodologies [9]. For example, [8] presents a system that discretizes the workspace into sectors and assigns sensors once an object enters a given sector (without *a priori* knowledge of its trajectory). In [10], attention-based behavior is used to reduce a multi-target sensing problem to a single-target one. More recent approaches often use agent-based strategies (e.g., [12]). One must note, however, that these works consider only fixed-geometry objects, and do not address effects specific to time-varying-geometry objects.

B. Form and Action Recognition

Recognition of an articulated form is a difficult task for an automated system, often requiring active sensing [13]. The earliest efforts in form recognition focused on reconstructing a model for an *a priori* unknown object [14]. Human gait recognition emerged as a central topic, which advocated that it is possible to uniquely identify an individual based on their gait. As such, algorithms using key-point markers to distinguish the current form of a human, given a single image, were developed [15]. Hierarchical approaches showed that decision-level data fusion could facilitate human identification (e.g., [16]).

Recent pertinent research has been reported in papers that classify approaches into three general categories: *template matching*, *semantic approaches*, and *statistical approaches* [7], [17], [18], [19]. These action-recognition methods, however, do not consider sensor-pose choice as part of the methodology, nor viewing angle [13]. Viewpoints exhibit non-uniform importance – when recognizing the actions of articulated objects, they are differentiated by both pose and time [20]. If the system has input data on certain parts of the overall model, views containing images of the other, missing parts become relatively more important. Furthermore, subject self-occlusion increases viewpoint differentiation due to sensor pose – it increases error only for some viewpoints [3]. Any articulated object may self-occlude, but all occlusions will change over time. As such, there is a need for continuous surveillance [2].

Recent vision-based methodologies targeting time-varying-geometry objects have been mostly adapted versions of reconfiguration algorithms developed for fixed-geometry objects (e.g., [21]). This paper presents a novel, multi-camera active-vision system, proposed to improve form-recognition performance by selecting near-optimal viewpoints. The proposed method seeks to maximize the visibility of the time-varying-geometry target, hereafter referred to as Object of Interest (OoIs), moving in a cluttered, dynamic environment.

II. Problem Formulation

The problem at hand is recognition of the actions of one or more OoIs that may move on *a priori* unknown paths. Furthermore, the workspace may be cluttered with multiple, dynamic obstacles, also moving on *a priori* unknown paths. This necessitates path prediction for both the OoI and obstacles, which inherently introduces a margin of error. Complex articulated objects may also self-occlude, and they are subject to non-uniform importance of external viewpoints. Thus, the chosen methodology must be robust and operate in real-time.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_72,

This paper proposes using '*the success rate in recognizing current target form and action*' as a metric of performance. It, thus, depends on the quantity and quality of sensor data collected. This in turn depends on visibility of the OoI to a given sensor (i.e., camera), which is captured herein by a visibility metric, V. The visibility metric for the i^{th} camera at the j^{th} demand instant, t_j, expressed as $V_i^j = f_i^j(\boldsymbol{p}_{S_i}^j)$, is taken as a function of $\boldsymbol{p}_{S_i}^j$, the pose of the i^{th} sensor, S_i, at the j^{th} instant. Pose is defined as a 6D vector $[x \;\; y \;\; z \;\; \varphi \;\; \psi \;\; \theta]$ representing position, (x, y, z), and orientation, (φ, ψ, θ). This metric actually depends on the current form and pose of the OoI, and the current poses of the obstacles and sensors. Only the camera poses can be directly controlled by the system and, thus, considered variables. For a system sensing a time-varying-geometry object, this paper proposes a global formulation of the reconfiguration problem with n_{sens} sensors, n_{obs} obstacles, and with prediction over the time horizon ending at the m^{th} demand instant:

```
For each demand instant, tj, j=1 to m:
    For each sensor, Si, i=1 to nsens, solve:
```

Given: $\boldsymbol{p}_{S_i}^0, \boldsymbol{p}_{OoI}^0, \boldsymbol{u}^0, \boldsymbol{p}_{obs_k}^0$; $k = 1$, to n_{obs},

Maximize: $$Pr = g\left(V_i^l\right);\ l = 1 \text{ to } j, \tag{1}$$

Subject to: $$\boldsymbol{p}_{S_i}^l \in P_i, \tag{2}$$

$$\boldsymbol{p}_{S_i}^l \in A_i^l, \tag{3}$$

$$V_i^l \geq V_{min};\ l = 1 \text{ to } j, \tag{4}$$

```
    End of loop.
Continue while: tproc < tmax,
```

Above, $\boldsymbol{p}_{OoI}^j$ is the pose of the OoI at the j^{th} demand instant, $\boldsymbol{p}_{obs_k}^j$ is the pose of the k^{th} obstacle at the j^{th} demand instant, $\boldsymbol{u}^j$ is the feature vector of the OoI at the j^{th} demand instant, P_i is the discretized set of feasible camera poses for the i^{th} camera, A_i^j is the discretized set of achievable camera poses for the i^{th} camera at the j^{th} demand instant, V_{min} refers to a user-defined threshold of minimum visibility, t_{proc} is the time spent processing data, and t_{max} is the maximum time limit for the selection of a final pose set. The two sets of feasible and achievable poses, P_i and A_i^j respectively, are defined, where a pose is feasible if it is within the physical motion limits of the sensor. The achievable poses are, thus, a strict subset of the feasible poses, representing those that can be reached in the time remaining before the *deadline*.

This problem formulation uses a performance objective function, Pr, which in turn depends on the visibility metric of each camera at all demand instants on the horizon. An implementation capturing this problem formulation seeks to maximize performance by first maximizing the visibility of the OoI at the immediate future demand instant, [t^1], for all cameras. If sufficient time remains, the system seeks to maximize expected visibility at [t^1 and t^2], then, [t^1, t^2, and t^3], and so on. As such, a higher overall metric value may be achieved at later demand instants, possibly at the expense of near-future visibility – this trade-off can be controlled by adjusting the minimum desired visibility.

System Calibration

In keeping with the qualitative goal of robustness, a formal camera calibration method is necessary to manage this key source of error [22]. One can define an optimization problem that captures the goal of a generic camera calibration method used to track the 2D projection of a 3D point as:

```
For each sensor, Si, i=1 to nsens, solve:
```

Minimize: $$E = \left\| \boldsymbol{x}_{projected_{Si}} - \boldsymbol{x}_{pixel_{Si}} \right\| \tag{5}$$

```
End of loop.
```

The vector $\boldsymbol{x}_{projected_{Si}}$ is the projection of the true 3D point through the camera model (not yet specified) into 2D pixel coordinates. Similarly, $\boldsymbol{x}_{pixel_{Si}}$ contains the actual pixel coordinates of point in question, and E is simply the straight-line distance between the two. In a real-world system, $\boldsymbol{x}_{pixel_{Si}}$ will be given by a 2D feature recognition or tracking algorithm and, thus, will be subject to additional noise.

The recovered parameters in this minimization come from $\boldsymbol{x}_{projected_{Si}}$, which depends on the chosen camera model. For most vision systems, these are intrinsic parameters, for example focal lengths or principal point coordinates, and possibly distortion coefficients to account for various lens distortions [23]. However, this neglects camera motion – a second, critical source of error in a system with reconfigurable sensors. As such, the proposed methodology includes a complete system calibration method that solves the minimization problem (5) while controlling constant and scaled parameters of error in rotation/translation of the camera.

III. PROPOSED GENERALIZED METHODOLOGY

The proposed sensing-system reconfiguration methodology consists of multiple agents: *sensor agents* (one per physical sensor), a *referee agent,* a *central planner agent,* a *prediction agent,* and a *form/action recognition agent.*

A. Sensor Agents

The exact configuration of the physical sensors (in terms of number and composition of the sensor set) can be determined through any number of established methods (e.g., [6]). A sensor agent encapsulates all movement capabilities of the physical sensor, managing its current 6-dof pose over time:

$$t_d = t_1 - t_0, \tag{6}$$

$$\boldsymbol{L}_{min} < \boldsymbol{x}_1 < \boldsymbol{L}_{max}, \tag{7}$$

$$\boldsymbol{x}_{L-} < \boldsymbol{x}_1 < \boldsymbol{x}_{L+}, \tag{8}$$

$$\boldsymbol{x}_{L-} = f(\boldsymbol{x}_0, t_d),\ \boldsymbol{x}_{L+} = f(\boldsymbol{x}_0, t_d), \tag{9}$$

Above, for a given sensor agent in motion, $\boldsymbol{x}_0$ is the initial position, $\boldsymbol{x}_1$ is the final position, t_0 is the initial time, t_1 is the final time, t_d is the total time between the demand instants, and $L_{min/max}$ are the outer limits of the motion axis. Similarly, $\boldsymbol{x}_{L-}$ and $\boldsymbol{x}_{L+}$ are the minimum and maximum poses achievable, respectively, given the capabilities of the sensor, the current pose, and the time remaining. Together, these equations capture the limited physical motion of the sensor. Equation (9) defines an arbitrary function for determining the achievable limits. A similar set of equations can be used to determine the rotational limits in terms of angular velocity and acceleration.

Each sensor agent is also responsible for evaluating its visibility metric, at current and future poses. One can model all objects in the environment as elliptical cylinders, bounding the maximum space the object can occupy. A clipped projection plane can be established, allowing the agent to project all objects onto this plane. The metric is, thus, given by:

$$V = W_{area} f(T_1 \dots T_a) + W_{dist} f(f, c_{obj}) + W_{angle} f(\phi) \tag{10}$$

In Equation (10), W_{area}, W_{dist}, and W_{angle} are the weighting constants for the metrics of visible OoI area, zoom factor (size of OoI in the camera view), and angular distance from a view with the OoI centered in the image, respectively. The vectors $\boldsymbol{T}_1$ to $\boldsymbol{T}_a$ define a polygon bounding the 2D projected area of the OoI, taken at the sensor pose being evaluated. The vector $\boldsymbol{f}$ is the focal point of the camera, and $\boldsymbol{c}_{obj}$ is the center of the OoI, such that the second term function is proportional to the distance between the camera and subject (size of the subject in the captured image). Finally, ϕ is the angle between the camera focal center and the subject center.

A sensor agent is also responsible for maintaining calibration parameters for its associated camera and movement stages. In previous sensing-system methodologies, a traditional camera calibration technique, such as Tsai's camera calibration method, was used with a de-rotation and de-translation matrix to return the rotated camera reference frame to the original frame the sensor was calibrated in. Such methods are still adaptations of methods developed for static cameras, and as such can be a significant source of error. Our methodology includes a novel system calibration method.

The implemented methodology used in the experiments is based on the 'Plumb Bob' distortion model (which is the basis for some modern techniques, such as [23]) given in [24], and Tsai's calibration method [22]. The core projection from world coordinates, $\boldsymbol{X}_W$, is taken from [22]:

$$\boldsymbol{X}_C = \boldsymbol{R} X_W + T \tag{11}$$

$$\boldsymbol{X}_W = \begin{bmatrix} x_{target} - x_{stage} \\ y_{target} - y_{stage} \\ z_{target} - z_{stage} \end{bmatrix} \tag{12}$$

Here, $\boldsymbol{X}_C = [x_c \quad y_c \quad z_c]^T$ are the coordinates of the point in camera-frame coordinates, the matrix $\boldsymbol{R}$ contains the rotation model used (as a function of the current stage rotations), and $\boldsymbol{T}$ is a constant translation vector. The values x_{target}, $y_{target,}$ and z_{target} (the target point) and x_{stage} , y_{stage} , and z_{stage} (camera center point given by translation stage) are given in world coordinates. The key improvement from the proposed methodology is a calibrated, flexible mapping from known stage coordinates to world coordinates, by replacing (12) with:

$$\boldsymbol{x}_w = \begin{bmatrix} x_{target_m} + kx_{1,1} x_{stage_m} + kx_{1,2} x^2_{stage_m} + \dots + kx_{1,I} x^I_{stage_m} + kx_{2,1} y_{stage_m} + \dots + kx_{2,J} y^J_{stage_m} + \dots + kx_{3,K} z^K_{stage_m} + x_{con} \\ y_{target_m} + ky_{1,1} x_{stage_m} + ky_{1,2} x^2_{stage_m} + \dots + ky_{1,I} x^I_{stage_m} + ky_{2,1} y_{stage_m} + \dots + ky_{2,J} y^J_{stage_m} + \dots + ky_{3,K} z^K_{stage_m} + y_{con} \\ z_{target_m} + kz_{1,1} x_{stage_m} + kz_{1,2} x^2_{stage_m} + \dots + kz_{1,I} x^I_{stage_m} + kz_{2,1} y_{stage_m} + \dots + kz_{2,J} y^J_{stage_m} + \dots + kz_{3,K} z^K_{stage_m} + z_{con} \end{bmatrix} \tag{15}$$

Here, $kx_{1,1..3,I}$ $ky_{1,1..3,J}$ $kz_{1,1..3,K}$ are translation function coefficients for the mapping function for x_w, y_w, and z_w respectively. The values x_{con}, y_{con}, z_{con} are constant offsets. This mapping accounts for non-linearity within the motion stage itself, and for errors in absolute positioning (including off-angle stage positioning and other subtle errors). A similar model can be used for the stage rotations (although it can be difficult to calibrate for, depending on the model $\boldsymbol{R}$ used).

Lastly, a method of tracking the key articulation points must be implemented within the sensor agent. For simplicity, the experiments presented herein will use an established tracking methodology to produce 2D pixel estimates of each joint visible to a given sensor agent. Local PCA (Principal Component Analysis) will be used to initially identify articulation points, which will be tracked by an optical flow algorithm. If the sensor agent detects that a point has lost tracking, the PCA step will be repeated to re-establish a lock. Each point will also use a corresponding Kalman Filter (KF).

B. Central Planning Agent

The use of a central planning agent is proposed to accept the sorted visibility evaluations from each sensor agent, as well as a list of the discretized achievable poses, in order to generate camera assignments and select their final poses for the demand instant at hand. A simple set of rules can be used:

- Only the M highest visibility cameras are assigned.
- Cameras with a visibility metric less than a minimum, V_{min}, at all poses, are not assigned: they await assignment in anticipation of potentially optimal future viewpoints.
- For assigned cameras, a weighted sum of metrics is evaluated, which includes feedback from the form-prediction agent on which sub-parts of the OoI are not currently well represented in the dataset.

Assuming that each articulation point is actively tracked by each sensor agent, one can identify if each is visible in the associated camera field-of-view. Sensor agents will capture new frames, and track points, as quickly as possible. The planner, on the other hand, will operate to select new poses only at a specific rate, given by the choice of demand instants. Each sensor agent attempts to maintain an open communication link with every other sensor agent and the central planner. The system operates under the following rules:

- Each sensor agent broadcasts (to all connected agents) the detected locations of any visible points with each new frame, along with a timestamp.
- If an agent receives newer estimated pixel coordinates, as seen by another camera, it rebroadcasts this information to each connected agent that does not already have a copy.

In order to facilitate the above transactions, the clocks of each agent are assumed to be already synchronized. In this manner, each agent retains the latest detected pixel coordinates of every point from the perspective of every sensor agent. The central planner maintains the Kalman filter associated with each articulation point (see *Section A*). This scheme maximizes the likelihood that the central planner will have the most current position estimates possible available to it. More importantly, the method is robust to communication link failures (a temporary latency increase constitutes a link failure when demand instant spacing is close). The time overhead for guaranteed delivery methods cannot be afforded.

Finally, when a feature vector request is made, reconstruction begins. The goal is to reconstruct the feature vector at the exact time the request was made, t_{req}:

For each articulation point, n = 1..n_{points}:
    For each sensor, s = 1..n_{sens}:
        If $t_{obs_{Sensor\_s}} \geq t_{req}$, use the equivalent observation with $t_{obs_{Sensor\_s}} = t_{req}$ for $\mathbf{x}_p$, or an interpolation if none are equal.
        Otherwise, do one of the following:
            1. Predict $\mathbf{x}_p$ using the Kalman track, only if $t_{obs_{Sensor\_s}} - t_{req} \leq t_{max}$.
            2. Wait for t_{wait} until new data arrives
        End Choice.
    End For.
End For.

Here, $t_{obs_{Sensor\_s}}$ is the timestamp of the latest observation of the point by *Sensor s*, t_{max} is the maximum time difference to allow prediction over, and t_{wait} is a minimum amount of time to wait for. World coordinates are recovered through a robust line intersection method, such as the iteratively re-weighted least-squares method (robust *m*-estimator) used in the experiments. The intersection must respect the inverse of the camera model.

C. Pose-Prediction Agent

This agent predicts the future poses of the OoI, and all obstacles, from historical data. A number of established options exist, such as the Kalman Filter and its variants.

D. Referee Agent

The goal of this agent is to ensure globally desirable behavior, by defining a set of global rules for the system to follow. Typically, such rules are highly application specific. For example, a rule could be defined to guarantee the assignment of a minimum number of cameras at each instant to the surveillance of the OoI.

E. Form/Action Recognition Agent

The form/action recognition agent is not directly responsible for maintaining estimates of the feature vector of the OoI. Instead, it requests a current feature vector from the central planner agent. A model-based method is used to recognize the current OoI form and action.

IV. EXPERIMENTS

The goal of this paper is to demonstrate that the proposed methodology is capable of sensing a single moving target in a dynamic environment with multiple, potentially moving obstacles.

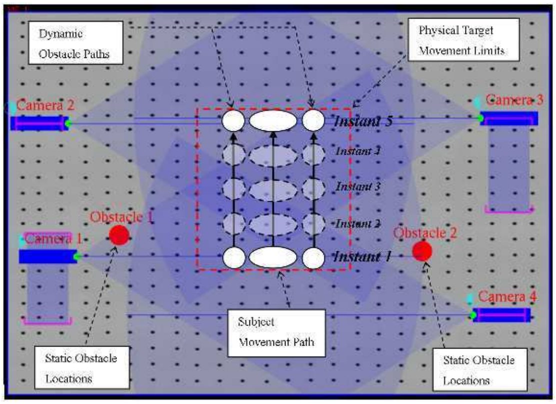


Figure 1 – Annotated experimental set-up overview, including motion paths.

A. Experimental Set-up

The experimental set-up consists of a single-target, dynamic environment, with four active cameras. Each camera is mounted on a rotational stage, while two cameras have an additional linear stage. The OoI is mounted on an *X-Y* stage in the center of the workspace. Two cylindrical obstacles are also present. This layout and paths for all objects are shown in Figure 1. A simple, articulated human model is used – the model contains 14 dof, and is at a 1:6 scale to an average human subject. Model poses are duplicated manually, and as such all experiments are quasi-static. Form recognition was achieved by a model-based algorithm [3], using an optical-flow/robust-intersection method to track articulation points.

In all experiments, the OoI maintained a constant velocity of 0.1 m/s on a straight-line path through the workspace center. For the 'slow' reconfigurable system, the maximum camera velocity was 0.045 m/s, with a maximum acceleration of 0.9 m/s^2. For the 'fast' system, maximum camera velocity is 0.45 m/s, with a maximum acceleration of 9 m/s^2. All initial camera poses are identical, and were selected assuming the subject is uniformly likely to be at any location within the workspace. Static obstacles were placed roughly at the center-point of the OoI motion, one on each side of its path. For the dynamic obstacle trials, the obstacles move alongside the OoI.

For the reconfiguration problem, a fixed assignment of three cameras to the nearest demand instant, t_1, and one camera to the next instant, t_2, is used. Demand instant spacing is 1 s. Furthermore, each sensor agent uses a weighted combination of three sub-metrics to rank poses: the area of the target bounding cylinder that is visible, the angle to the target, and the distance to the target. An upper limit of 0.125 was selected for the overall error metric. Any trial/instant with an error value above this limit is considered not to yield a positive match. This value was determined through statistical analysis of multiple previous runs, resulting in at least 95% true positive matches for un-rejected frames, and at most 2% false negatives [2]. The error metric itself is a weighted sum of the mean error in the recovered feature vector to the true OoI library form, and two additional terms proportional to the number of un-recovered, and unusable key-points, respectively [2]. From past feasibility studies, the error metric should be considered as having an additive noise component (mostly caused by the manual pose duplication process), resulting in a ±0.01 bound (approximately) on any given error metric value. Error bars have been excluded from the graphs for visibility.

B. Experimental Results

A total of 9 trials were carried out, for all combinations of [*no obstacles*, *static obstacles*, *dynamic obstacles*] and [*no reconfiguration*, *slow reconfiguration*, *fast reconfiguration*].

Case A: No Obstacles

The first set of trials, presented in Figure 2, is designed to show a baseline comparison with no obstacles present.

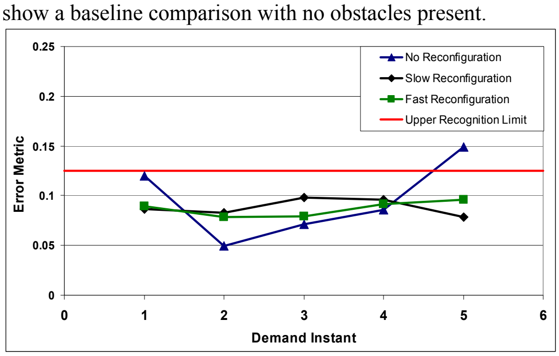


Figure 2 – Experimental results for no obstacles.

From the initial trial with no system reconfiguration, one can note that at Instants 2-4, the subject form is recognized with relatively low error. Here, most sensors have at least a partial view of the subject. For Instants 1 and 5, the target is out of the FOV of at least one camera (Cameras 2 and 3 for Instant 1, and 1 and 4 for Instant 5), and recognition quality degrades. The addition of system reconfiguration in the slow and fast reconfiguration trials significantly reduces the error metric for these two demand instants, by keeping the subject in view of all cameras.

As expected, the system selects nearly identical sensor poses, in the absence of any obstacles, for slow and fast reconfiguration. The primary source of error is the manual subject-form reproduction, which accounts for most of the (minimal) differences in error metric. Due to global optimality issues, the slow and fast reconfiguration trials have the system select a near-optimal pose for instant 1. However, this choice results in a set of achievable poses for instant 2 that does not include the original static camera poses (which happen to result in a lower error metric). Note that by instant 3, the fast system is able to correct this problem, and selects poses almost identical to the static case. The slow system takes until instant 4 – this is the price of reconfiguration: in some cases global performance is improved, but at the cost of short-term performance. Note that the slow and fast systems select basically identical sensor poses in instant 5, but the additive noise in the recognition process obscures this fact.

Case B: Static Obstacles

In the case of static obstacles, Figure 3, trials with no system reconfiguration correctly recognize the subject form for Instants 2 and 3, but not 1 and 5. Now, the error metric for Instant 4 has also significantly increased – the obstacles occlude several key views (primarily for Cameras 1 and 4). When system reconfiguration is used, slow or fast, the negative effect of the static obstacles is neutralized. The system selects near-optimal viewpoints that are not occluded by the static obstacles. As such, these trials clearly show that reconfiguration can improve recognition performance by removing the effect of one or more static obstacles.

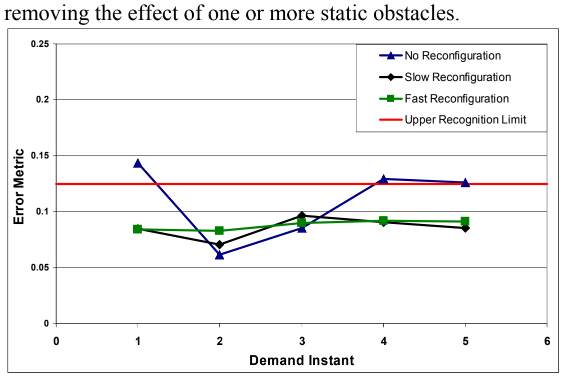


Figure 3 – Experimental results for static obstacles.

Case C: Dynamic Obstacles

The final set of trials, Figure 4, presents the most challenging reconfiguration problem – dynamic obstacles, with a high proportion of partial and complete occlusions.

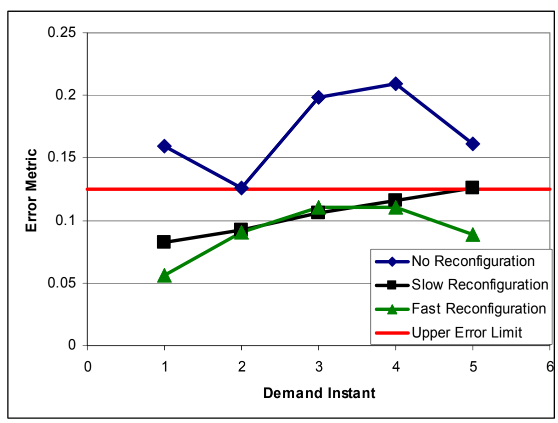


Figure 4 – Experimental results for dynamic obstacles.

From Figure 4, the subject form is considered not recovered at any instant when the cameras are static. However, the use of sensing-system reconfiguration has clearly improved performance – all instants are now recognized.

V. Conclusions

In this paper, a novel sensing-system reconfiguration method, designed for sensing the actions of articulated human subjects, was presented. The method uses an agent-based approach, selecting near-optimal viewpoints along a prediction horizon, to remove the effects of environment and OoI-specific factors. Controlled system-reconfiguration experiments demonstrated a tangible improvement in *real-world* form-recognition performance over static sensors. In particular, the system was able to significantly reduce the error metric of a recovered form, even in the presence of multiple, dynamic obstacles.

Acknowledgments

We acknowledge the financial support of the Natural Science and Engineering Research Council of Canada (NSERCC).

References

[1] K.A. Tarabanis, P.K. Allen, and R.Y. Tsai, "A Survey of Sensor Planning in Computer Vision," *IEEE Transactions on Robotics and Automation*, vol. 11, no. 1, pp 86-104, Feb. 1995.

[2] M. Mackay and B. Benhabib, "A Multi-Camera Active-Vision System for Dynamic Form Recognition," *International Joint Conference on Computer, Information, and Systems Sciences, and Engineering (CISSE2007),* Dec. 2007.

[3] M. Mackay, R. G. Fenton, and B. Benhabib, "Time-Varying-Geometry Object Surveillance Using a Multi-Camera Active-Vision System," *International Journal on Smart Sensing and Intelligent Systems,* vol. 1, no. 3, pp. 679-704, Sept. 2008.

[4] J. Miura and K. Ikeuchi, "Task-Oriented Generation of Visual Sensing Strategies in Assembly Tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 2, pp. 126-138, Feb. 1998.

[5] X. Gu, M. M. Marefat, and F. W. Ciarallo, "A Robust Approach for Sensor Placement in Automated Vision Dimensional Inspection," *Proc. of the 1999 IEEE Intl. Conf. on Robotics and Automation,* Detroit, MI, pp. 2602-2607, May 1999.

[6] C.K. Cowan and P.D. Kovesik, "Automated Sensor Placement from Vision Task Requirements," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 3, pp. 407-416, May 1988.

[7] T. Urano, T. Matsui, T. Nakata, and H. Mizoguchi, "Human Pose Recognition by Memory-Based Hierarchical Feature Matching," *Proc. of IEEE International Conference on Systems, Man, and Cybernetics*, pp. 6412-6416, The Hague, Netherlands, 2004.

[8] B. Horling, R. Vincent, J. Shen, R. Becker, and K. Rawlings, "V. Lesser: SPT Distributed Sensor Network for Real Time Tracking," Technical Report 00-49, University of Massachusetts, Amherst, MA, 2000.

[9] R. Niepold, S. Sakane, and Y. Shirai, "Vision Sensor Set-up Planning for a Hand-Eye System Using Environmental Models," *Proc. of the Society of Instrument and Control Engineers of Japan*, vol. 7, no. 1, pp. 1037-1040, Hiroshima, Japan, 1987.

[10] T. Darrel, I. Essa, and A. Pentland, "Attention-Driven Expression and Gesture Analysis in an Interactive Environment," *International Workshop on Automatic Face and Gesture Recognition*, pp. 782-790, Zurich, Switzerland, 1995.

[11] A. Bakhtari, M. Mackay, and B. Benhabib, "Active-Vision for the Autonomous Surveillance of Dynamic, Multi-Object Environments," *Journal of Intelligent and Robotic Systems*, 2008.

[12] A. Bakhtari and B. Benhabib, "An Active Vision System for Multi-Target Surveillance in Dynamic Environments," *IEEE Transactions on Systems, Man, and Cybernetic*s, vol. 37, no. 1, pp. 190-198, 2007.

[13] S. Yu, D. Tan, and T. Tan, "A Framework for Evaluating the Effect of View Angle, Clothing, and Carrying Condition on Gait Recognition," *Proc. of Int. Conf. on Pattern Recognition,* pp. 441-444, Hong Kong, China, 2006.

[14] S.Y. Chen and Y.F. Li, "Vision Sensor Planning for 3-D Model Acquisition," *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, vol. 35, no. 5, pp. 894-904, Oct. 2005.

[15] R. Chellappa, A.K. Roy-Chowdhury, and S K. Zhou, "*Recognition of Humans and Their Activities Using Video,*" San Rafael, CA: Morgan & Claypool Pub., pp. 53-92, 2005.

[16] A. Kale, A.K. Roy-Chowdhury, and R. Chellappa, "Fusion of Gait and Face for Human Identification," *Proc. of ICASSP04,* pp. 901-904, Montreal, Canada, 2004.

[17] M. Dimitrijevic, V. Lepetit and P. Fua, "Human Body Pose Recognition Using Spatio-Temporal Templates," *ICCV workshop on Modeling People and Human Interaction*, pp. 127-139, Beijing, China, Oct. 2005.

[18] D. Cunado, M. S. Nixon, and J. Carter, "Automatic Extraction and Description of Human Gait Models for Recognition Purposes," *Computer Vision and Image Understanding*, vol. 90, pp. 1-41, 2003.

[19] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, and H. Zhang. "Human Gait Recognition with Matrix Representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 896-903, July 2006.

[20] R. Bodor, P. Schrater, and N. Papanikolopoulos, "Multi-Camera Positioning to Optimize Task Observability," *Proc. of IEEE Conf. on Advanced Video and Signal Based Surveillance*, pp. 552-557, 2005.

[21] L. Hodge and M. Kamel, "An Agent-Based Approach to Multi-sensor Coordination," *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, vol. 33, no. 5, pp. 648-662, Sept. 2003.

[22] R. Tsai, "A Versatile Camera Calibration Technique for High Accuracy 3D Machine Vision Metrology using Off-the-shelf TV Cameras and Lenses," *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323-344, 1987.

[23] J. Heikkila, and O. Silven, "A Four-step Camera Calibration Procedure with Implicit Image Correction," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1106-1112, 1997.

[24] D. C. Brown, "Decentering Distortion of Lenses," Photometric Engineering, pp. 444-462, Vol. 32, No. 3, 1966.

Effects of Online MCQ Tests on Student Learning

B. Issac and S.M. Jacob
Swinburne University of Technology (Sarawak Campus),
Jalan Simpang Tiga, Sarawak, MALAYSIA.

***Abstract*-Getting students to understand the concepts of a technically sound subject – Business Data Communication and Networks, done by non-technical (i.e. business) students was the challenge. A set of Multiple Choice Questions (MCQ), around 40 to 50 of them per chapter were posted on the subject Blackboard as online MCQ tests that the students could take at their own pace, after a chapter is finished. They were encouraged to do it as a part of their learning exercise though no marks were given for it. Once they complete the test, they could see the answers and the mistakes they made, through immediate feedbacks. There were eight students who participated. The MCQ sets were posted online in two sessions. In session 1, MCQs for the first four chapters were posted. Later, their understanding of concepts was checked through a mid-semester test, covering these four chapters. In the second session, MCQs for the next seven chapters were posted and their understanding was tested through a final examination, covering these seven chapters. The output graphs of the assessment scores show generally, that the students who did the MCQ tests well, performed well in mid-semester test and final exam too. Those who did badly comparatively in MCQs, fared poorly in mid-semester test and final exam as well. Some exceptions were there, where some student(s) who did a bit of hard work through MCQs, did not perform that well (but not very badly too) in the test or exam. A regression analysis is done between MCQ test results and the mid-semester test and final exam. Quantitative and qualitative analyses done on a brief survey conducted with the students confirmed the effects of the observed learning process, through the use of online MCQs. This could very well be extended to large classes.**

I. INTRODUCTION

Educationists are becoming more aware of the need for assessment tools which could be used as tools for learning rather than merely for assessment or grading [1]. The authors believe that when teaching students, making students understand the concepts behind the topic is more important than the act of teaching itself. This has formed the main objective of this paper as we use the online learning management system to implement chapter-wise multiple choice questions to enhance learning. A traditional multiple choice question (MCQs) is one in which a student chooses one answer from the choices or alternatives supplied (normally five choices like A, B, C, D and E). Basically, MCQs consists of [2] the question (that can be called – stem), the choices provided after the stem (that can be called – options or alternatives), the correct answer in the list of options (that can be called – key) and distractors which are the incorrect answers in the list of options as shown in figure 1.

The three key features of a test are its content validity, clarity and reliability. A test's content validity is determined by how well it covers and tackles the range of knowledge, skills, and abilities that students should have gained in the time duration (weeks or months) covered by the test. Test clarity points to the measure of the standard of questions in terms of quality that covers the subject details. Test reliability relies on the consistency of grading and the steps to ensure non-discrimination among students of different performance levels [3]. All these issues have been considered, in the design and implementation of the online MCQ tests that forms the paper.

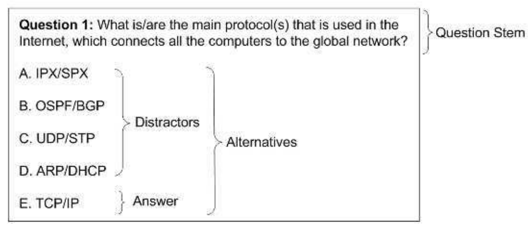


Figure 1. Structure of MCQ question and alternatives

The authors are investigating both quantitatively and qualitatively the merits of using Multiple Choice Questions in an online environment as a tool for learning, rather than just for assessment sake. This paper is organized as follows. The section 2 covers the related studies and works, section 3 depicts in detail the analysis of assessment marks, section 4 explains on the quantitative results of the online survey, section 5 deals with a short discussion on the qualitative comments for the online survey and section 6 is the final discussion and conclusion.

II. RELATED STUDIES AND WORK

Atkinson and O'Connor [4], in their paper investigate the effectiveness of Multiple Choice Questions (MCQs) in a computer environment as a means of absorbing information. It compares two computerized methods having different degrees of interactivity to a more traditional paper-based method. The increase in learning is quantified using pre-tests and post-tests. The study involved pupils in two second level schools in the Dublin area and also investigated their attitudes towards such use of computers for this purpose. Kwen and Cheng [5] in their paper discuss on the usage of two tier reflective multiple choice questions to cater to creative thinking. The creative test taker has no avenue to explain his/her reason for choosing a particular option as his/her choice of the answer key. The study involves the use of a modified test instrument (a "reflective" or

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_73, © Springer Science+Business Media B.V. 2010

"two-tier" MCQ) which comprises items from the traditional MCQ with an added second tier which is essentially an open-ended segment which requires pupils to explain the thinking behind their choice of the answer key. The study sample comprises two primary level classes, primary four and six in Singapore schooling system. Ramesh, Sidhu and Watugala [6] in their paper discusses the development of an unconventional, yet an effective method for administering short and simple MCQs as a means of assessing student learning in a focus area of the subject matter. A MCQs test was designed, wherein the five choices (n = 5) of answer to a question is tagged with different numerical numbers based on the multiplication of numeral (n + 1) and not by the alphabets A, B, C, D and E as per normal practice. At the end of the test, the candidate will total the selected choices for questions and the test result can then be obtained by inputting the total choice value in a simple spreadsheet programme. The spreadsheet programme was written specifically to decode that number and indicate the number of correct answers. Nicol [1] in his paper does an analysis that shows the different ways in which MCQs can be used to support the development of learner self-regulation. The framework and principles are offered as a way of helping teachers design the use of MCQs in their courses and of evaluating their effectiveness in supporting the development of learner autonomy. A key message from this analysis is that the power of MCQs (to enhance learning) is not increased merely by better test construction. Power is also achieved by manipulating the context within which these tests are used. Williams [7] in his paper reflects on the ongoing debate surrounding the usefulness (or otherwise) of multiple choice questions (MCQ) as an assessment instrument. The context is a graduate school of business in Australia where an experiment was conducted to investigate the use of assertion-reason questions (ARQ), a sophisticated form of MCQ that aims to encourage higher-order thinking on the part of the student. It builds on the work of Connelly [8] which produced a quantitative analysis of the use of ARQ testing in two economics course units in a flexibly-delivered Master of Business Administration (MBA) program. Harper [9] discusses that the Multiple-choice tests should be implemented with rigorous attention to quality by following 'good practice' to avoid the pitfalls in their design. He asserts contrary to some suggestions, such MCQ tests can be used to assess a variety of outcomes including all of the competences in Bloom's taxonomy and the desire to assess the use of knowledge does not require more complex question structures such as Extended Matching Sets Questions (EMSQ)'s. MCQ tests are a useful assessment tool, with a simple design, easy implementation (with appropriate staff development) and with clarity of focus.

III. Analysis Of Assessment Marks

The primary objective of this research study is to see how online MCQ tests can augment student learning. This was checked quantitatively through (1) the comparative descriptive statistics of the scores of the course assessments – mid-semester test and final exam and the MCQ test scores; (2) the multiple linear regression model tested, taking the final exam score as the dependent variable, and the mid semester marks, average MCQ (session 1) marks, average MCQ (session 2) marks as the independent variables.. The authors planned to make learning more interactive digitally and to get the students to read their text books well in order to find the answers. The study was based on 8 students of a 14-week long course in Computer Science (namely, Business Data Communication and Networks), within a university in Malaysia. The assessments of the course included a mid-semester test (in week 8) and a final exam (in week 15). A series of eleven MCQ tests were enabled on the Blackboard Learning System by the course instructor for the different chapters of the course as in figure 2. The set of MCQ tests, around 40 of them per chapter, covering all nooks and corners of a chapter, could be taken by the students at their own pace, once a chapter is finished during the lecture. The marks of these tests did not contribute to the internal assessment so as to reduce their stress of 'doing a test every week'. But the students were highly encouraged to do it as a part of their learning exercise. They need not do it in one go, but could save their answers and continue later. Once they complete the test, they could see the answers and the mistakes they made, through immediate feedbacks as in figure 3. The first five of the eleven MCQ tests happened before the mid-semester test, and the last six of the MCQ tests were done by the students after the mid semester test.

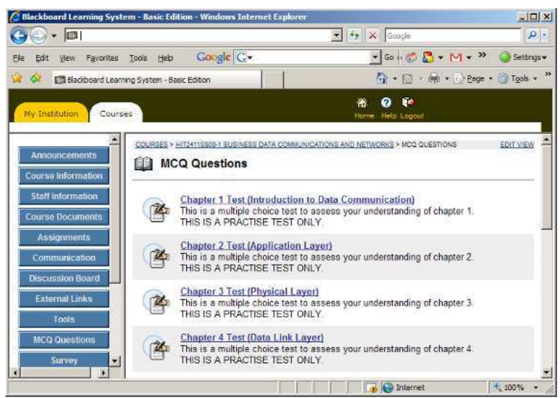


Figure 2. Blackboard links of the online MCQ Tests

Figure 4 shows the results of mid-semester tests to be following the MCQ performance (in MCQ 1-4) approximately. This is indicative of student learning happening as they engage in online MCQ exercises. Figure 5 does a student-by-student comparison of the mid-semester test results and the average MCQ scores. As shown in Figure 4, the mid semester test marks vary from 50-90 and is proportionate to the range of the MCQ test marks which varied from 30-90. The means and standard deviations of the mid semester test marks and the average MCQ (average of MCQ1 to MCQ4) test marks, in table 1, show the proximity in the values between the two. But the sample is neither varied nor big enough to confirm a clear correlation between the two sets of scores.

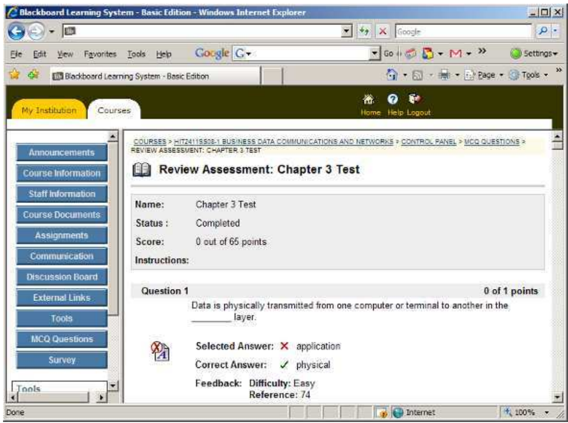


Figure 3. MCQ feedbacks after the test, on Blackboard

TABLE I. DESCRIPTIVE STATISTICS FOR THE MID SEMESTER AND AVERAGE MCQ (1-4) TEST MARKS

Marks	Mean	SD	N
Mid Semester	74.00	15.91	8
Average MCQ (session 1)	70.53	19.23	8

Looking at the individual student scores from figure 4, student 1, 2, 3, 6 and 7 generally performed well in the MCQ tests (there are some cases where the student did badly for one MCQ test only) and the mid semester test. Figure 5 shows the comparative trend of the progress of the two sets of marks. There's a general progress across both sets of marks, though there are clear exceptions like student 2 (who did MCQ 4 only badly).

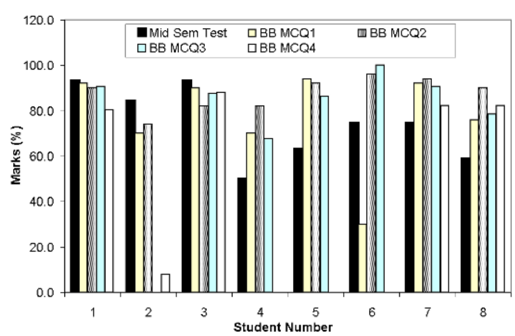


Figure 4. The scores for mid-semester test and individual MCQ tests (for 1-4) in session 1

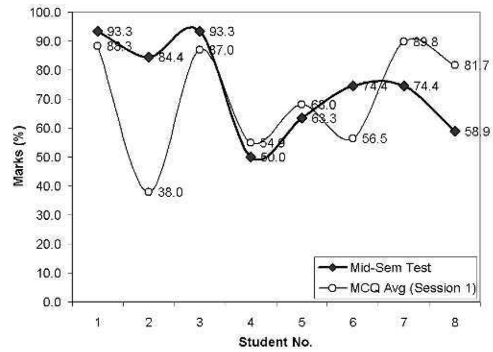


Figure 5. The scores for mid-semester test and the average of MCQ test marks (for 1-4) in session 1

A similar set of graphs have been done for MCQ tests 5-11 and the final exam score in figures 6 and 7. As observed previously, the final exam scores are consistently higher than the MCQ test scores for each student, or have a close relationship to MCQ performance, when a student has taken the MCQ tests well. Exception to this would be students 5 and 6. Table 2 confirms the superiority of Final Exam marks over the average MCQ (session 2) marks, with mean of final exam clearly greater than the corresponding mean of the average MCQ (session 2) marks. There is noticed a moderate correlation between the two sets of marks (Pearson correlation, $r = 0.5$, $r^2 = 0.25$), which points to the fact that final exam scores and average MCQ (session 2) marks were positively influencing each other.

TABLE II. DESCRIPTIVE STATISTICS FOR THE MID SEMESTER AND AVERAGE MCQ (6-11) TEST MARKS

Marks	Mean	SD	N
Final Exam	76.4	8.9	8
Average MCQ (session 2)	51.9	24.0	8

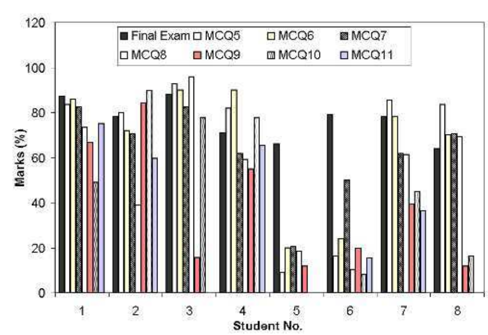


Figure 6. The scores for final exam and individual MCQ tests (for 5-11) in session 2

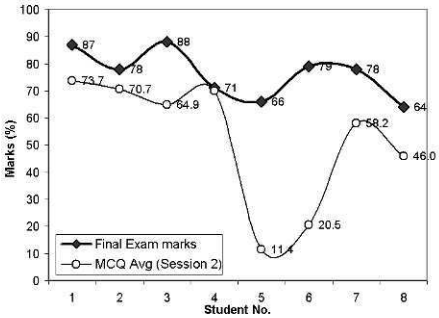


Figure 7. The scores for final exam and the average of MCQ tests (for 5-11) in session 2

A multiple linear regression model was tested using SPSS software, taking the final exam score as the dependent variable, and the mid semester test marks, average MCQ (session 1) marks, i.e. avg_MCQ1 and average MCQ (session 2) marks, i.e. avg_MCQ2 as the predictor variables. The regression model showed a strong positive correlation among the dependent and predictor variables (Pearson correlation, $r = 0.908$, $r^2 = 0.824$). The ANOVA Table 3 shows the progression of introducing the three predictor variables, as model 1 to 3. Model 3 is the one that has been quoted here. A significant relationship is observed between the three predictor variables and the dependent variable in the model ($F = 6.232$, $p \approx 0.05$) at 10% significance level.

TABLE III. ANOVA TABLE OF THE REGRESSION MODEL

ANOVA[d]

Model		Sum of Squares	Degrees of freedom	Mean Square	F	Sig.
1	Regression	134.977	1	134.977	1.952	0.212[a]
	Residual	414.898	6	69.150		
	Total	549.875	7			
2	Regression	157.020	2	78.510	0.999	0.431[b]
	Residual	392.855	5	78.571		
	Total	549.875	7			
3	Regression	452.963	3	150.988	6.232	0.055[c]
	Residual	96.912	4	24.228		
	Total	549.875	7			

a. Predictors: (Constant), avg_MCQ2
b. Predictors: (Constant), avg_MCQ2, avg_MCQ1
c. Predictors: (Constant), avg_MCQ2, avg_MCQ1, mid-sem test
d. Dependent Variable: final exam

Table 4 shows the contribution of the different predictor variables (using the regression coefficients) to the model, separately for model 1-3. The average MCQ (session 1) marks is not contributing significantly to the model ($t = 0.076$, $p = 0.942$) compared to the mid semester marks ($t = 3.495$, $p = 0.025$), average MCQ (session 2) ($t = 0.928$, $p = 0.406$) marks. But on the overall, the multiple regression model explains the total contribution of the three coursework marks- average MCQ (session 1), average MCQ (session 2) and the mid semester marks, to the final exam. This indirectly points to the complimentary role of the MCQs to the building up of the students' final exam results.

TABLE IV. TABLE OF THE REGRESSION COEFFICIENTS

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	66.867	7.413		9.020	0.000
	avg_MCQ2	0.183	0.131	0.495	1.397	0.212
2	(Constant)	60.687	14.092		4.307	0.008
	avg_MCQ2	0.176	0.140	0.477	1.256	0.265
	avg_MCQ1	0.093	0.175	0.201	0.530	0.619
3	(Constant)	38.527	10.072		3.825	0.019
	avg_MCQ2	0.077	0.083	0.208	0.928	0.406
	avg_MCQ1	0.007	0.100	0.016	0.074	0.945
	mid-sem test	0.450	0.129	0.809	3.495	0.025

a. Dependent Variable: final exam

IV. QUANTITATIVE ANALYSIS OF THE SURVEY

An online survey was done at the end of the 14-week course, to get feedbacks from the students who participated in the learning through MCQ tests, so that the authors would know the impact of the adopted approach. The quantitative results of that survey are given below, along with the posted questions.

1. Do you agree with the statement – "The MCQ Test (non-assessed) experience was a great learning style, where we could read through chapters and understand the concepts well".

The opinion on the learning through MCQ online tests as in figure 8 were greatly positive (around 85%, though 15% was neutral), as the authors noted that dynamic and interactive applications stimulates the younger generation compared to the static data on books without any interactivity.

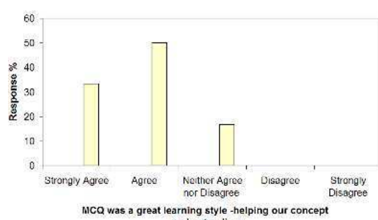


Figure 8. Feedbacks on online MCQ as a great learning style

2. MCQ experience for this subject, enhanced/increased my reading habits.

The response for MCQ tests enhancing the reading habits as in figure 9 was positive (greater than 65%, though over 30% or more were neutral). One of the objectives of the MCQ tests was to make the students search for answer from the textbook by reading it.

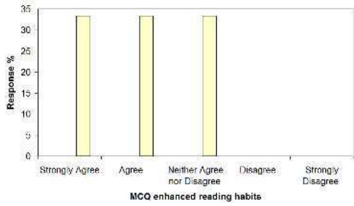


Figure 9. Feedbacks on online MCQ enhancing the reading habits

3. As the subject is quite technical, MCQ approach is very good to make learning a fun (as we can do it in our own pace.

The general picture that the authors observed was that the students were quite eager to do the online tests, as there was no compulsion and that they could do it at their own pace. They were told that, if they could do the MCQ tests well, they could do the written tests also well. The response as seen in figure 10 was 100% positive.

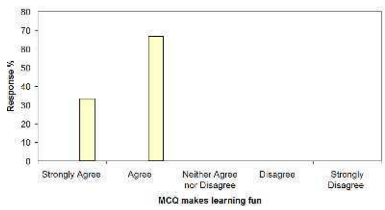


Figure 10. Feedbacks on fun learning through online MCQ

4. The major benefit from MCQ approach is the immediate feedback that I could get after completion, so as to correct my errors.

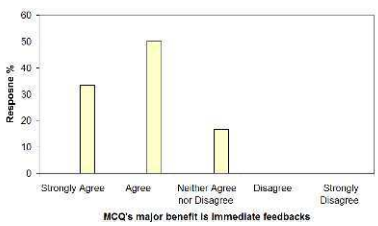


Figure 11. Feedbacks on the immediate feedbacks in online MCQ

For paper based exam, one of the disadvantages is the time to manually mark it and to give the feedbacks to the students. There can be thus a delay of few days to a week or more depending on the number of students doing that subject. With a online MCQ test, the feedback time is shortened. Once they complete the test, they would get immediate feedbacks with corrections and page number of that question's content in the text book. The response was over 80% positive and around 15% neutral as in figure 11.

5. As the MCQ approach is "digital" (i.e. done through a computer) and not on paper (hard copy), it makes it more attractive for student learners.

The students' liking for digital technology and making it a learning medium was over 80% positive and 15% or more neutral, as in figure 12. This was quite expected, as young learners have a general fascination toward anything digital.

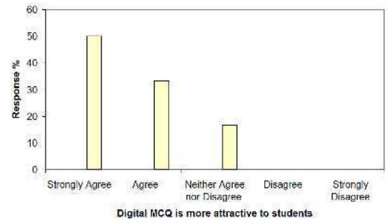


Figure 12. Feedbacks on the digital nature of online MCQ

6. Answer/Select all that apply. The most attractive aspect(s) about MCQ learning is – (1) It is good in giving immediate feedbacks, upon completion; (2) It can be done at my own pace (as I can save it and continue later); (3) It can be done in a relaxed environment (as it is non-assessed), but still helps in learning; (4) It makes me to search for answers (thus improve my reading) and compels me to learn.

All the stated advantages have a positive response of 50% to 70%, which reiterates the points discussed before on why online MCQ is advantageous.

7. Answer/Select all that apply. MCQ approach for learning is NOT advisable, because – (1) It is a boring approach and make me sick of learning; (2) It is time-waster and it never improves reading or learning!; (3) It is not very effective in understanding of concepts; (4) Not applicable, as I don't agree to this statement.

Only around 15% says that learning in not very effective through online MCQ tests and around 70% says that they disagree with the statement – "MCQ approach for learning is NOT advisable".

V. Qualitative Analysis Of The Survey

The students were asked for qualitative feedbacks on their learning experience through MCQ tests for one semester (14 weeks) and they are listed as follows.

- I support it, because it gives us specific grounds on where to concentrate our studies and with it being online we never loose it. The only thing is sometimes the Internet is down and we do not have access to it, but all in all it's a good idea to improve learning.
- It is indeed nice but we need to use a lot of time to complete it. If possible, don't put too many MCQ questions.
- It really could help me study through the chapter. I think it would be better to count this in for the assessment, to add some mark for us. But do not use time limit for completion as I enjoy working on it at my own pace.
- This MCQ test makes me read again the book and it can touch on the area that I miss out while doing the regular tutorial. I like MCQ more than the tutorial as it is really fun if I can get the answer exactly with no guessing.

VI. Final Discussion and Conclusion

Formation of good quality MCQ test is vital for the right kind of learning to happen in students. Some basic tips of forming good quality multiple choice questions [10] were found noteworthy by the authors.

1. A vertical and logical ordering of alternatives is essential so that the answers can be located without problems.
2. Avoid double negatives, tough English language constructions so that the students can easily understand want is being asked. Avoid long sentences and repetitive words in question stem and alternatives.
3. Have only one response to be the correct answer and avoid ambiguity.
4. The distracters should be effective so that it should confuse an incompetent student.
5. Always give clear indications to choose the best answer, rather than the correct answer, to avoid any misunderstandings later.
6. Use negatives in questions sparingly and if necessary emphasize words in question stem to help increase the clarity of what is being asked, while such emphasize in alternatives should be avoided.
7. Use 'none of the above' alternatives with caution and avoid it, if possible. Also avoid alternatives that are indirect like 'both C and D', 'all but B' etc.
8. Make the initial questions relatively easier, so that the student can be gradually led to questions of higher difficulty.
9. Avoid testing multiple concepts in one question. If the information given to students is fairly large, then testing on multiple concepts through multiple questions looks good.
10. Five alternatives (A) to (E) or (i) to (v) are always recommended.
11. Some of the strategies to counter students who try to make guess work are as follows: make all alternatives equally probable, make the longest alternative correct only one-fifth of the time, lessen use (or do not use) of alternatives like 'none of the above' or 'all of the above' and if used, make appear about one-fifth of the time.
12. Do not create the MCQ test in haste, as it would be wise to spend time on it to achieve the desired quality. Always check for ambiguities with a trusted colleague (wherever possible) before releasing the test.

The authors paid considerable attention in ensuring that the online multiple choice questions were generally of a good quality in relation to the facts stated above. From our study, a moderate positive correlation was observed between the two sets of marks – final exam scores and average MCQ (session 2) marks, which points to the fact that final exam scores and average MCQ (session 2) marks were positively influencing each other. The multiple regression model tested between final exam score as the dependent variable, and the mid semester test marks, average MCQ (session 1) marks and average MCQ (session 2) marks as predictor variables showed a strong positive correlation among the dependent and predictor variables. Implementing the online MCQs for each chapter of a subject or course and asking students to participate in it, would definitely enhance student reading and learning, as they need to search for answers from book. On a personal note, we found that the student's text book is full of markings done by highlighters, which at least shows that they are reading their book. Nevertheless, to get all the MCQs done for all chapters, can be a great challenge, if we are to do from the scratch. With the relevance of education in the world we are living in, with increasing student numbers in courses, alternative methods for effective assessment are all too welcome. Also there is a great demand from all quarters to release the results of the examination as soon as possible [11]. To tackle this demanding scenario, computers are increasingly utilized to mark test scripts that is implemented as multiple choice questions [12]. Once done, the MCQ tests could be reused and the marking along with feedbacks could all happen automatically and efficiently.

References

[1] D. Nicol, "E-assessment by design: using multiple-choice tests to good effect," *Journal of Further and Higher Education*, vol. 31(1), pp. 53–64, 2007.

[2] V. L. Clegg and W. E. Cashin, "Improving multiple-choice tests," Idea Paper No. 16, Center for Faculty Evaluation and Development, Kansas State University, 1986.

[3] J. P. Sevenair and A. R. Burkett, "Difficulty and discrimination of multiple choice questions: A counterintuitive result," *Journal of Chemistry Education*, vol. 61, 441-442, 1988.

[4] R. Atkinson and R. O'Connor, "The Application of Multiple Choice Questions in a Computer Environment as a Method of Learning rather than of Assessment," MCE Working Papers 1999, Dublin City University, Ireland, pp.1-11, 1999.

[5] B. H. Kwen and A. K. Cheng, "Using two tier reflective multiple choice questions to cater to creative thinking," AARE 2005 International Education Research Conference, pp.1-11, 2005.

[6] S. Ramesh, S. M. Sidhu and G. K. Watugala, "Exploring the Potential of Multiple Choice Questions in Computer-Based Assessment of Student Learning," *Malaysian Online Journal of Instructional Technology*, vol. 2(1), pp.1-15, 2005.

[7] J. B. Williams, "Assertion-reason multiple-choice testing as a tool for deep learning: A qualitative analysis," *Assessment and Evaluation in Higher Education*, vol. 31(3), pp.287-301, 2006.

[8] L. B. Connelly, "Assertion-reason assessment in formative and summative tests: results from two graduate case studies," in: R. Ottewill, E. Borredon, L. Falque, B. Macfarlane & A. Wall (Eds) Educational innovation in economics and business VIII: pedagogy, technology and innovation (Dordrecht, Kluwer Academic Publishers), 2004, pp. 359–378.

[9] R. Harper, "Multiple-choice Questions – A Reprieve," *Bioscience Education eJournal*, 2003. [Online]: http://bio.ltsn.ac.uk/journal/vol2/beej-2-6.htm

[10] R. A. Dewey, Writing Multiple Choice Items which Require Comprehension,[Online]: http://www.psywww.com/selfquiz/aboutq.htm, Retrieved on August 15, 2008.

[11] P. Hosie and T. Mazzarol, "Using technology for the competitive delivery of education services," *Journal of Computer Assisted Learning*, vol.15, pp.118-128, 1999.

[12] P. Williams, "The learning web: the development, implementation and evaluation of internet-based undergraduate materials for the teaching of key skills," *Active Learning in Higher Education*, vol. 3, pp.40-53, 2002.

Knowledge Assessment – Practical Example in Testing

Jan GENCI
Department of Computers and Informatics, FEEI, Technical University of Kosice, Slovakia
genci@tuke.sk

Abstract-**Knowledge assessment is inseparable part of current e-learning technologies. It can be used for self-assessment of students to give them feedback about their progress in a study or for an intermediate or final grading for tutors. However, knowledge tests are not developed with the adequate care. Author's experience in the area of knowledge assessment led to a confidence the "unstructured" testing is usually used in this process. It means, that many of knowledge tests are not designed to reveal the reached level of knowledge. Moreover, testing suites are reviewed very seldom regarding their validity and items correlation. This paper presents experiences gained during the design and implementation of a specific software focused on teaching several principles of the Unix.-like operating systems. The structure of the specific assignment follows the Bloom's taxonomy of educational objectives.**

I. INTRODUCTION

E-learning proves itself as a viable mode of study, which can support classical approach and in some cases can be the only mode for education. Knowledge assessment is an inseparable part of the pedagogy and the current e-learning technologies. It is usually used for self-assessment of students to give them feedback about their progress in a study or for the intermediate or final grading for tutors. However, knowledge tests are very often not developed with the adequate care. Author's self-experience in the area of knowledge assessment led to a confidence the "unstructured" testing is usually used in this process. It means that testing suites are not designed to reveal the reached level of knowledge according to some taxonomy of cognitive aims. Moreover, testing suites are very seldom reviewed regarding their validity and items correlation.

Another problem is the increasing number of the enrolled students. Most of the Slovak universities accepted obligation stemmed from Bologna declaration and introduced three level study structure for university education (undergraduate, graduate and doctorate studies). This transformation requires restructuring of educational plans/programs. Due to the parallel courses the number of students significantly increases.

That was the situation regarding the "Operating Systems" subject taught at our department during the winter term in academic year 2006/2007. Students were assigned two assignments – the first one regarding various aspects of FAT (File Allocation Table) file system and the second one regarding the usage of the Unix/Linux system calls. Both of the assignments did not require students to present overall knowledge of the materials taught in the course but only the small part of it. Moreover, the assignments were on various level of difficulty.

During the past years, there was a necessity to interview every student at the end of the semester in order to classify the level of knowledge s/he reached (credit assignments and exams). However, assessment of knowledge based on interview results may be quite subjective. The situation may be even worse if the interview is based on different conditions (in the case of credits - assignments assigned to students during the semester).

To solve the presented drawbacks of the assignments we decided to develop two complex assignments. The main goal was to verify deeply students' knowledge regarding FAT file system and Unix/Linux system calls.

This contribution describes the approach we chose for knowledge assessment in the area of the FAT file system. Presented approach allowed us to assess about 500 assignments during the semester in parallel for students in the second and third term. Nowadays we teach about 300 regular students and number of students has increasing tendency.

II. TAXONOMY OF EDUCATIONAL OBJECTIVES

Taxonomies of educational objectives are well known in the pedagogical theory. They try to specify path a learner has to pass in order to achieve a certain level of knowledge. The higher level of knowledge is always built on the lower level knowledge. It means, learner who reached a specific level of knowledge is expected to be able to demonstrate proficiency in any of the lower levels of taxonomy.

The most known taxonomy is Bloom's taxonomy [1] and its revisions [2]. In Slovakia, Niemierko's taxonomy [3] is quite often cited and sometimes stated as more suitable for building cognitive (knowledge) tests. Fuller [10] refers to some other taxonomies (Tollingerova's, Bespalko's and SOLO), which are widely used in the design of courses and development of assessment tests. Specifics about computer science related courses and suitability of presented taxonomies are also discussed and presented in some examples. As a result, new taxonomy for computer science is presented in the work of Fuller[10]

In the next analysis we will focus on the following "basic" taxonomies.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_74, © Springer Science+Business Media B.V. 2010

Bloom's taxonomy

According to this taxonomy, educational objectives are divided into three domains:

- Cognitive;
- Affective;
- Psychomotor.

Only cognitive domain of the Bloom's taxonomy presents interest for the knowledge assessment in our approach. Cognitive domain according Bloom's taxonomy specifies six levels [4]:

- Knowledge: recognize or recall information;
- Comprehension: demonstrate that the student has sufficient understanding to organize and arrange material mentally;
- Application: a question that asks a student to apply previously learned information to reach an answer;
- Analysis: higher order questions that require students to think critically and in depth;
- Synthesis: higher order question that asks the student to perform original and creative thinking;
- Evaluation: a higher level question that does not have a single correct answer.

Niemierko's taxonomy

Niemierko's taxonomy consists of four levels [5]:

- Remembering – similar to knowledge in Bloom's taxonomy;
- Understanding – similar to comprehension in Bloom's taxonomy;
- Specific transfer – application of acquired information according presented patterns;
- Non-specific transfer – creative application of acquired information.

It was mentioned already that taxonomy of educational objectives is accepted and applied in pedagogy, but its application in the knowledge test design and implementation is not so straightforward. Moreover, positioning of testing item to particular level can be very specific in technology or computer science, as discussed in the work of Fuller [10].

III. Knowledge Assessment

Term "didactic test" is often used in Slovak and Czech pedagogical community. It usually denotes process of examination designated for fair evaluation of level of mastering of a given topic. Term "achievement test" is more frequently used in the English speaking countries.

Regardless of what term is used, this type of tests is considered as tool of systematic measure of results of learning.

It is accepted, that didactic/achievement test has to be prepared according to specified rules by team of experts. We have to define what is measured by the test, target group, administration of the test, way of test evaluation and assumed conclusions according achieved results [6], [7], [11].

Test can contain closed and opened test items. Closed items are items for which several answers are prepared and participants have to decide what a right answer is (i.e. true false, one from many choice, many from-many choice, sorting, matching etc.). Opened items are items for which answers are not specified and participants are required to fill/submit right answer.

Bloom and his team already in 1956 [1] observed that 95% of testing items requested students to recall information only. The situation is not change dramatically to present days. Problem lays, to our opinion, in current testing tools, which do not support testing items for higher levels of cognitive taxonomy. Such testing items require open answers and their evaluation can be extremely difficult.

IV. FAT File System Knowledge Evaluation

To assess knowledge regarding functionality of FAT file system we requested students to design and implement a computer program for transfer of files from hosting operating system to image of FAT file system (exact copy of file system stored in the regular file or emulated FAT file system).

Required knowledge

FAT file system comprises several data structures and several basic and advanced algorithms. Basic data structures:

- Bios Parameter Block (BPB) – base data about file system;
- File Allocation Table (FAT) – contains data about allocation of clusters to file;
- Root Directory (RootDir) – root of hierarchical file system;
- Area of clusters - data area, where raw data from files and directories are stored;
- Directory Entry structure (DirEntry) – item , where data about certain file are stored (name, attributes, size of file, date etc);
- Directory structure – array of DirEntries, which contains data about all files in the directory.

Basic algorithms

- Reading of BPB and calculation of basic information about file system (starting point of FAT, RootDir, area of clusters, number of DirEntries in RootDir);
- FAT operations (read item corresponding to cluster, read next FAT item; find free FAT item)
- Reading/writing one cluster of data;
- Finding file in the directory by scanning all valid DirEntries;

Complex algorithms

- Finding relevant directory by recursive search of directory tree;
- Finding chain of clusters (FAT items chain) comprising data area of the file;
- Read file content;
- Create DirEntry;
- Write file content.

Mapping of specific knowledge to the Bloom's taxonomy levels is as follows.

Students are asked to pass all 6 levels within solving the assignment:

- knowledge – students have to learn structure (BPB, FAT, RootDir, cluster organization, DirEntry etc) of all versions of FAT file systems (FAT12, FAT16, FAT32) and basic operational algorithms (reading of BPB, FAT item, cluster data based on its number etc);
- comprehension – students have to demonstrate their understanding of the purpose of every part of the FAT file-system, relationships between its parts, their mutual dependency;
- application – students are required to apply acquired knowledge in the process of (re)design and (re)implementation of basic algorithms which operate on FAT file system structures;
- analysis – copying file from host file system to image of FAT file system require detailed analysis of operation of FAT file system (reading BPB; extraction of file system parameters; localization of FAT table, root directory and first data block – cluster; processing of directory; recursive localization of cluster corresponding to relevant directory; etc);
- synthesis – based on performed analysis and basic building blocks prepared in the early phases students are required to "synthesize" computer program for specified purpose;
- evaluation – as was stated above, this level of taxonomy requires questions that do not have a single correct answer. Implementation of complex computer program never has single solution. It can be carried out by many different ways (what is used for plagiarism detection).

Some authors require for the evaluation level of Bloom's taxonomy even higher knowledge – to be able to compare effectiveness of two or more programs performing the same task. In this case we should position our assignment to synthesis level.

Similar specification of the required taxonomy level can be presented for the Niemierko's taxonomy. Just to prove it – we require students to creatively apply acquired knowledge regarding FAT file systems what means the highest level of Niemierko's taxonomy - Non-specific transfer.

Evaluation of assignments

Students submit their solution in the form of the C/C++ source code program to assignment tracking and evaluation system. Several requirements have to be fulfilled by students during coding and submitting of their work in order to successfully evaluate the solution by the system. Students are not limited in the number of submission attempts.

Each student's submission is compiled and carefully tested in the secured environment. Assignment submitted by student is executed several times. Any single execution is performed with different parameters – image of FAT file system and specification of files which have to be 'read from' or 'write to' emulated FAT file system. Detailed protocol from the assignment evaluation process is provided to student. Moreover, students are able to include their own outputs at any place in the source code.

Source code submitted to the system can be considered as answer to open testing item (question). Such type of item must be thoroughly analyzed and evaluated. In the case of the FAT file system assignment evaluation is based on the parallel "inverse" execution of required operations with "baseline" program prepared by the test item developer. Inverse execution means different things for 'read from' or 'write to' assignments.

In the case of 'read from' assignment, some files from the host file system are copied to emulated FAT file system ('written to' - that is a reason we call it "inverse" operation to 'read from' assignment). Such emulated FAT file system is provided to evaluated program together with possible wildcard specification of files, which should be extracted by student's program. After extraction, number of extracted files, their names and content are compared.

In the case of 'write to' assignment, preliminary emulated FAT file system is provided to evaluated program together with possible wildcard specification of files, which reside in the host file system. Student's program is required to place (write) relevant files to emulated FAT file system. After execution of student's program, FAT file system image is analyzed by extraction of relevant file ('read from' – again, that is a reason we call it "inverse" operation to 'write to' assignment). Number of extracted files, their names and content are compared.

V. Conclusion

Student, who submitted assignment and her/his assignment was evaluated successfully, was considered having appropriate knowledge regarding FAT file system and its operation.

However, anti-plagiarism feature, built in the assignment tracking and evaluation system discovered, that only 10% of students really designed and developed their assignments. The rest were plagiarism - source code was usually taken from colleagues, very often slightly modified. To be correct, it is necessary to say, that students were allowed to use "third party solutions", but they were required to understand it. Understanding was checked by asking questions about data structures, algorithms and/or source code itself. However, the most of the students identified by plagiarism detection software were not able to answer the questions.

Our experience show that presented approach requires very carefully prepared text of assignment. In the first version of assignment there were not exactly specified all necessary conditions. This problem was solved by forum, where students asked about ambiguities and where explanations were provided.

Based on the experience learned from running FAT knowledge evaluating assignment, we decided to split assignment to several simplified assignments in the future. We plan to assign basic algorithms tasks, corresponding to lower levels of Bloom’s taxonomy (levels 1-3) during the first half of the semester as a separate assignment(s). Later, during the second part of semester we plan to require students to solve more complex tasks, corresponding to upper levels of Bloom’s taxonomy (levels 4-6).

Presented assignment was introduced in the Operating Systems subject in an academic year 2003/2004 together with implementation of Assignment tracking and evaluation system [9]. As mentioned above, in academic year 2006/2007 we were able to manage students in two parallel courses, because of transition to three level study. During that period we served more than 500 students. In summary, we served more than 1000 students till the present time.

References

[1] Taxonomy of Educational Objectives: The Classification of Educational Goals; pp. 201-207; B. S. Bloom (Ed.) Susan Fauer Company, Inc. 1956.

[2] Anderson, L. W., & Krathwohl, D. R. (eds.) (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. New York: Longman.

[3] Niemierko B.: Taksonomia celow wychowania. In: Kwartalnik pedagogiczny, 24, 1979, č.2 s. 67-78. (in Polish)

[4] The Taxonomy Of Educational Objectives. http://www.humboldt.edu/~tha1/bloomtax.html

[5] Gazdikova V.: Zaklady distancneho elektronickeho vzdelavania. Pedagogicka fakulta Trnavskej univerzity. Trnava. 2003 (in Slovak)

[6] Lavicky T.: Tvorba a vyuzivanie skolskych testov. (Ucebny text pre PVPZ a PV). Metodicko-Pedagogicke centrum Presov. Year not stated. (in Slovak) http://www.mcpo.sk/downloads/Publikacie/PrirodPred/PPCHE200501.pdf

[7] Stoffová V., Psenakova I.: Computer knowledge testing in distance education. Proceedings of international conference UNINFOS 2000, Nitra, 10.-12.5.2000, SPU v Nitre, 2000, pp. 309-313, ISBN 80-7137-713-9 (in Slovak)

[8] Tothova D.: Communication between tutors and students. Teaching and Learning in Virtual Learning Environments for Water Management. Nitra : Slovenska polnohospodarska univerzita (SPU), . 2007. pp. 7-8. ISBN: 978-80-8069-916-1

[9] Urbancik M. Management, Automation and Checking of Semestral Works. Diploma work. DCI FEEI Technical University of Kosice, 2005 (in Slovak)

[10] Fuller U. et all: Developing a computer science-specific learning taxonomy. ACM SIGCSE Bulletin, Volume 39, Issue 4 (December 2007). pp 152-170.

[11] Samuelis Ladislav, Szabo Csaba, Havlice Zdenek: Teaching Java programming based on the pool of open source case studies, Proc. of the 7th International Conference on Applied Informatics Vol. 2, ICAI 2007, Eger, January 28-31, 2007, pp. 55-62.

View-Independent Face Recognition with RBF Gating in Mixture of Experts Method by Teacher-Directed Learning

Salman Khaleghian[1] s.khaleghian@seiau.ac.ir
Nina Taheri Makhsoos[2] ni_ta78@stu-mail.um.ac.ir
Reza Ebrahimpour[3,4] ebrahimpour@ipm.ir
Alireza Hajiany[4] Alireza.hajiany@gmail.com

[1]Science and Research Branch Azad University, Tehran, Iran
[2]Department of Computer Engineering, Ferdowsi University, Mashhad, Iran
[3]School of Cognitive Sciences, Institute for Studies in Theoretical Physics and Mathematics, Tehran, Iran
[4]Department of Electrical Engineering, Shahid Rajaee University, Tehran, Iran

Abstract- **The present study focuses on using a new model to perform view-independent human face recognition. A model based on mixture of experts is proposed, which uses teacher-directed learning method to force the experts to learn a predetermined partitioning of the input face space, using a Radial basis function neural network for Gating network. This way, each expert obtains expertise over faces of a same pose. Experimental results on the PIE dataset demonstrated the improved performance of our proposed model in comparison with ME in its conventional learning style in terms of higher recognition rate.**

I. INTRODUCTION

Computer face recognition has received tremendous amounts of attention during last decades. A challenging task is to achieve face recognition under the constraint that the face has only been previously observed from different angles. Various models in view-independent face recognition can be categorized into three classes of Multiview, 3D Model and View-Invariant methods. Earlier methods focused on constructing invariant features [1] or synthesizing a prototypical view (frontal view) after a 3D model is extracted from the input image. A recent survey of approaches to 3D face recognition is provided in [2]. Such methods work well for small rotation angles, but they fail when the angle is large, say 60°, causing some important features to be invisible.

Most proposed methods are based on using a number of multiview samples. It seems that, in these methods, the most direct way of recognition is by simply storing a sufficient number of different views associated with each face, and then comparing the unknown image with all these views. Some models of associative memories propose that the huge memory capacity of the brain may be used for such a direct approach to recognition [3, 4]. Although useful, especially for the recognition of highly familiar faces, this direct approach by itself is insufficient for recognition in general. The main reason is the problem of generalization, which is, recognizing a face under a novel viewing direction. An example of this multiview approach is the work of Beymer [5], which models faces with templates from 15 views, sampling different poses from the viewing sphere. The recognizer consists of two main stages, a geometrical alignment stage where the input is registered with the model views and a correlation stage for matching. The main limitations of these methods are the need for many different views per person in the database, dependence on lighting variations or facial expressions and the high computational cost, due to iterative searching involved.

In this paper, we propose a neural computational model for view-independent face recognition which is based on the so-called mixture of experts (ME) architecture and falls under the category of multiview based approaches. However, in our model, we use modified of ME in which MLPs instead of linear networks for experts are used, and is hereafter referred to as mixture of multilayer perceptron experts, the problem space is divided into several subspaces for the MLP experts, and then the outputs of experts are combined by a gating network. In order to improve gating network performance and consequently the whole network performance, we using Radial basis function (RBF) in gating network.

To evaluate the performance of our proposed model, we use PIE dataset in our experiments, and we use Principal Component Analysis, PCA, in the feature extraction phase. PCA is one of the most common methods of dimensionality reduction that is widely used in the literature of face recognition for feature extraction [13]. Our method differs in the classification part, which

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_75,

is uses a TDL in the learning phase of Mixture of Experts. The performance of our method is compared to that of conventional ME and ME with MLP experts with Teacher-Directed learning.

The remainder of this paper is organized as follows. Section II, briefly described ME. It is followed by the description of our proposed model by details in Section III. Section IV. Section V presents experimental results and comparisons with previously published approach to the same problem. Section VI, finally draws conclusion and summarizes.

II. MIXTURE OF EXPERTS

From a computational point of view, according to the principle of divide and conquer, a complex computational task is solved by dividing it into a number of computationally simple tasks and then combining the solutions to those tasks. In supervised learning, computational simplicity is achieved by distributing the learning task among a number of experts, which in turn divides the input space into a set of subspaces. The combination of experts is said to constitute a combination of classifiers.

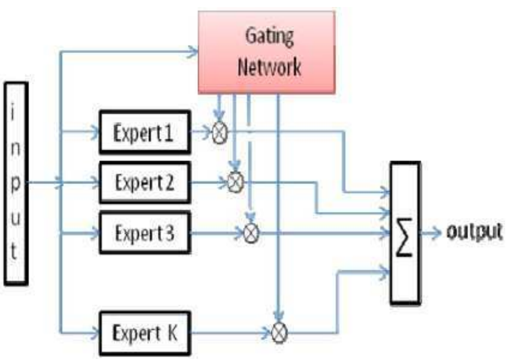


Figure 1. Block diagram of Committee Machine based on Mixture of Experts (the scalar outputs of the experts are mediated by a gating network).

Mixture of experts is one the most famous methods in the category of dynamic structures of combining classifiers, in which the input signal is directly involved in actuating the mechanism that integrates the outputs of the individual experts into an overall output [7]. Consider a modular neural network in which the learning process proceeds by fusing self–organized and supervised forms of learning as shown in Fig. 1. The experts are technically performing supervised learning in that their individual outputs are combined to model the desired response. There is, however, a sense in which the experts are also performing self–organized learning; that is they self–organize to find a good partitioning of the input space so that each expert does well at modeling its own subspace, and as a whole group they model the input space well. The learning algorithm of the mixture structure is described in [8].

For improve the performance of the expert networks devise a modified version of ME in which each expert is an MLP, instead of linear networks [9, 10]; In addition, we use a RBF network in gating network, which it is well-known that the RBF Networks overcome some of the MLP problems by relying on a rapid training phase, and presenting systematic low responses to input patterns that have fallen into regions of the input space where there are no training samples. Such characteristics and the intrinsic simplicity of these networks render RBF classifiers an interesting alternative to classifiers based on other neural models [11].

III. MIXTURE OF MLP EXPERTS AND RBF GATING NETWORK

We use MLPs instead of linear networks or experts in the informing the ME structure. In order to match the gating by RBF network and experts by MLP networks, the learning algorithm is corrected by using an estimation of the posterior probability of the generation of the desired output by each expert. Using this new learning method, the MLP expert networks' weights are updated on the basis of those estimations and this procedure is repeated for the training data set. The learning procedure is described in [12], and is briefly described in the following paragraphs.

Each expert is a one-hidden-layer MLP, that computes an output vector O_i as a function of the input stimuli vector x and a set of parameters such as weights of hidden and output layer and a sigmoid function as the activation function. It is assumed that each expert specializes in a different area of the face space. The gating by RBF network assigns a weight g_i to each of the experts' outputs, O_i.

The gating network determines the g_i as a function of the input vector x and a set of parameters such as weights of the hidden layer, the output layer and a sigmoid function as the activation function. The g_i can be interpreted as estimates of the prior probability that expert i can generate the desired output y. The gating network is composed of two layers: the first layer is an RBF network, and the second layer is a softmax nonlinear operator as the gating network's output. The gating network computes O_g, which is the output of the RBF layer of the gating network, then applies softmax function to get:

$$g_i = \frac{\exp(O_{gi})}{\sum_{j=1}^{N} \exp(O_{gi})} \qquad i = 1,2,...,5 \qquad (1)$$

So the g_i is nonnegative and sum to 1. The final mixed output of the entire network is:

$$O_T = \sum_i O_i g_i \qquad i = 1,2,...,5 \qquad (2)$$

The "normalized" exponential transformation of Eq. (1) may be viewed as a multi-input generalization of the logistic function. It preserves the rank order of its input values, and is a differentiable generalization of the "winner-takes-all" operation of picking the maximum value, so referred to as softmax.

The weights of MLPs are learned using the back-propagation, BP, algorithm, in order to maximize the log likelihood of the training data given the parameters. Assuming that the probability density associated with each expert is Gaussian with identity covariance matrix; MLPs obtain the following online learning rules:

$$\Delta w_y = \mu_e h_i (y - O_i)(O_i(1 - O_i))Oh_i^T \qquad (3)$$

$$\Delta w_h = \mu_e h_i w_y^T (y - O_i)(O_i(1 - O_i)) \\ Oh_i (1 - Oh_i) \qquad (4)$$

$$\Delta w_{yg} = \mu_g (h - g)(O_g(1 - O_g))Oh_g^T \qquad (5)$$

$$\Delta w_{hg} = \mu_g w_{yg}^T (h - g)(O_g(1 - O_g)) \\ Oh_g (1 - Oh_g) x_i \qquad (6)$$

where μ_e and μ_g are learning rates for the experts and the gating network, respectively, Oh_i is the output of expert network's hidden layer, and hi is an estimate of the posterior probability that expert i can generate the desired output y:

$$h_i = \frac{g_i \exp(-\frac{1}{2}(y - O_i)^T (y - O_i))}{\sum_j g_j (-\frac{1}{2}(y - O_i)^T (y - O_i))} \qquad (7)$$

This can be thought of as a softmax function computed on the inverse of the sum squared error of each expert's output, smoothed by the gating network's current estimate of the prior probability that the input pattern was drawn from expert i's area of specialization. As the network's learning process progresses, the expert networks "compete" for each input pattern, while the gating network rewards the winner of each competition with stronger error feedback signals. Thus, over time, the gate partitions the face space in response to the expert's performance.

IV. The Proposed Model

The present work introduces a method to use RBF network for gating network in ME-based model. Also for method of learning we use Teacher-Directed Learning. In this section, first describe RBG gating for mixture of experts explained in Section 3 then describe the method of learning.

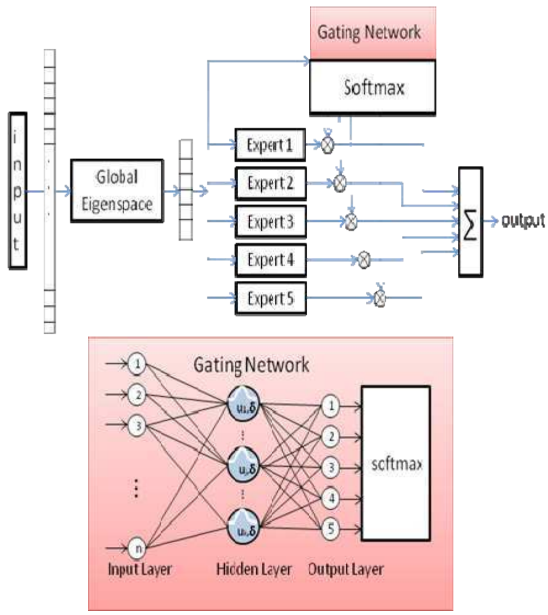


Figure 2. Sketch of the proposed model. Each input face image is first projected onto the global eigenspace and its output, which is a low dimensional representation of the input image, is presented to experts integrated in ME structure. In teacher-directed, experts are directed to specialize in predetermined views of faces. To recognize faces of novel views, in this model, the outputs of experts are combined to form the final output. The gating network shows by more details of RBF network.

A. RBF Gating Network

Individual experts are MLPs with one hidden layer. The gating network in our model is a radial basis function neural network (shown in Fig.2), constructed using the built-in Matlab function "newrb" as RBF network, in which network grows during training by adding neurons in hidden layer.

$$[net, tr] = newrb(P, T, gole, spread, MN, DF) \quad (8)$$

where P is input vectors, T is target class vectors, goal is Mean squared error goal, spread is Spread of radial basis functions, MN is Maximum number of neurons, DF is Number of neurons to add between displays.

Training phase in gating network included two levels. First, neurons of output layer are trained to minimize the total output error then when a neuron inserted gating network is retrained and error is approximated. The addition of neuron is run on until error value or maximum number of node (Eq.(8)). Spread shows in "newrb" by δ and centers chose near to the training dataset [13].

B. Teacher-Directed Learning in our Model

In the teacher directed learning method, teacher information is included in the training process, and it directs experts to produce appropriate outputs for their corresponding pose class of training samples. Note that in the testing phase teacher information on the pose of input faces is not present, thus, the network must infer the final states without that information. The detail of the training process is as follows: in the first step the training samples are projected onto the global eigenspace and are then fed into experts. Up to this stage, the training process is similar to that of self-directed partitioner ME.

To apply teacher-directed learning, we use the *Teacher* matrix, T, (Eq. (9)), in which T_j denotes the j^{th} column of T matrix.

$$T = \begin{bmatrix} 1 & 0.5 & 0 & 0 & 0 \\ 0.5 & 1 & 0.5 & 0 & 0 \\ 0 & 0.5 & 1 & 0.5 & 0 \\ 0 & 0 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0 & 0.5 & 1 \end{bmatrix} \quad (9)$$

Multiplying T_{ji} the i^{th} element of column T_j on h_i results in a new h which is nonzero for the j^{th} expert and its neighboring experts and is zero for all other experts. For right profile, right half profile, frontal, left half profile and left profile training samples, j is assigned the value of 1, 2, 3, 4 and 5, respectively. So expert 1 of the ME network shown in Fig. 2 serves as a +90°expert which also learns a portion of its neighboring subspace (+45°) , and expert 2 serves as a +45° expert which also learns a portion of its neighboring subspaces (+90° and 0°), and so on.

This way, with respect to the pose of each training sample there is one expert and its neighboring experts which have nonzero values of h. Zero h for experts mean that they are not allowed to update their weights (note that, according to Eqs. (3) and (4), the value of h defines the extent to which an expert's weights are updated, such that for $h = 0$ there is no weight updating and for $h = 1$ the maximum of calculated weight updates are applied). The weights of those experts with nonzero h are updated, proportional to their value of h, towards producing appropriate outputs for input training samples, in other words, they are directed to learn those input training samples which are of a same pose class. Now, some experts are directed towards a predetermined subspace by allowing just those experts to update their weights and keep others unchanged.

One might ask the reason for updating the weights of the neighboring experts. As a matter of face, we expect the model to combine the outputs of two or more experts to recognize faces of intermediate unseen views. For instance, for a face of +68.5° rotation, the gating network is expected to combine the outputs of +90° and +45° experts to form the final output. To endow the model the ability to interpolate between views, the experts should be able to learn a portion of their neighboring subspaces, in addition to their own division of face space; so they will be able generalize in recognizing faces of novel views. Thus, for each pose class, weights of the corresponding expert and the neighboring experts are updated.

V. EXPRIMENTAL RESULTS

In our experiments, the network's task is to recognize the faces of intermediate unseen views as individuals. We use a subset of the PIE database which consists of 10 identities with 9 different images of 9 different poses spaced evenly from -90° to +90° with 22.5° steps. Faces with ±45°, ±90° and 0° rotations are used to train and those with ±67.5° and ±22.5° rotations are used to test the networks Fig. 3. Fig. 4 shows examples of images used for training and testing the networks. To form the global eigenspace, face images in ±45°, ±90° and 0° rotations including the synthesized images are used. Therefore, we have 750 (10×5×15, 10 identities, each identity in 5 views and 15 images for each view) images, which by using the technique for PCA described in [14], we make a global eigenspace formed by the 50 eigenvectors of the training images covariance matrix.

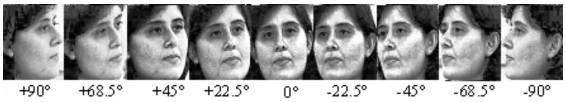


Figure 3. Examples of face images, taken from the PIE database.

Figure 4. Synthesizing new images. The single image at the top is the original, the images in the middle row are generated by changing the contrast and brightness of the original image and the images in the lower row are generated by applying Gaussian blur.

In all of the experiments learning algorithm is Teacher-Directed learning based on mixture of experts. As shown in Table I, our model based on ME with RBF gating network reveals better recognition rate than ME-based model with MLP gating network. It should be mentioned that the best result of this experiment for ME with MLP experts and gating networks is 84.10%, and for ME with MLP experts and RBF gating networks is 87.90%, with 20 hidden neurons in the former and 10 hidden neurons in the latter model, for the gating, respectively.

TABLE I

Recognition rates of different topologies of proposed models. In each row, for fixed values of hidden neurons of gating network, recognition rate which is the average of twenty training runs with different random initial weights is listed.

MLP Gating Network		RBF Gating Network	
Hidden layer Neuron	Percentage	Hidden layer Neuron	Percentage
5	69.03	5	68.89
10	75.6	10	85.98
15	80.85	15	83.05
20	84.02	20	81.65
25	81.75	25	82.81
30	80.65	30	83.09

We performed the experiment with Mixture of Linear Experts, Mixture of MLP Experts and gating networks, Mixture of MLP experts and RBF gating networks. The experiment was repeated for 20 times by randomly choosing different training and testing sets. Table 2 lists the details of the training parameters and the best structures, in terms of higher recognition rate, for the experts and gating network.

TABLE II

The details of the training parameters as well as the recognition rates of the ME, ME with MLP experts and gating network, ME with MLP experts and RBF gating networks on the test set.

Network Model	ME	ME with MLP experts	ME With RBF gating & MLP experts
Topology	5 Linear Experts	Experts: 50:65:20 Gating: 50:20:5	Experts: 50:65:20 Gating: 50:10:5
Learning rate	Experts: 0.01 Gating: 0.05	Experts: 0.01 Gating: 0.05	Experts: 0.01 δ:2.8
Max percentage	65.66	84.2	87.98

Table II summarizes the division of labor performed by each expert in both models over 20 times running with η_e = 0.01 and η_g= 0.05. The rows denote the recognition rate in different number of epoch of the five experts, first train experts then by MLP and RBF gating network tested. Spread in RBF gating network is increased incrementally from 0.1 in steps of 0.1 to 8.0. The best value based on experience is 2.8. Clearly, considering table III, for the same experts, irrespective of its topology, the experts reveal the same recognition rate. In RBF gating network is applied, there is strong improve performance in gating network.

VI. CONCLUSION

We proposed a model based on ME to achieve high performance view-independent face recognition. We compared two other ME-based models with our proposed one to demonstrate the superior performance. Experimental results supported our claim that there is better use RBF gating network in ME for view-independent face recognition when we use The basic idea was to employ another network for more performance to pose estimate in gating network with respect to linear or MLP network. By applying our model, the average recognition rate was increased.

ME appears to be a powerful tool to approach more robust view-independent face recognition models. In future work, we plan to explore more complicated methods of RBF network for use with experts or gating networks in ME structure.

TABLE III
Recognition rates of different gating network in the same expert. each row is shown with different number of epoch.

Number of epoch	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	MLP gating
25	31.25	41.25	38.75	40	30	77.25
40	33.25	41.25	42.5	42.5	30	80.25
70	36.25	40	45	42.5	31.25	83.01
100	36.25	45	45	42.5	35	**84**
140	36.25	42.25	46.6	42.5	33	83.25

Number of epoch	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	RBF gating
25	32.5	40.25	39.75	41	31	79.51
40	32.75	42.35	41.5	42.25	30.5	83.25
70	36.25	45	46.35	42.5	35	**86.48**
100	34.25	43.25	42.35	43.5	33.5	85.23
140	35.25	42.25	45.5	41.5	33	84.25

REFERENCES

[1] Wiskott, L., Fellous, J.M., and von der Malsburg, C.: Face Recognition by Elastic Bunch Graph Matching. IEEE Trans. Patt. Anal. Mach. Intell. **19** (1997) 775–779.

[2] Bowyer, K., Chang, P., and Flynn, A.: Survey of Approaches to Three-Dimensional Face Recognition. In: Proc. of the IEEE International Conference on Pattern Recognition, (2004) 358–361.

[3] Hopfield, J.J.: Neural Networks and Physical Systems with Emergent Collective Computational Abilities. In Proceedings of the National Academy of Sciences USA (1982) 2554-2558.

[4] Kohonen, T.: Associative Memories: A System Theoretic Approach, Berlin: Springer(1978)

[5] Beymer, D.J.: Face Recognition under Varying Pose. Technical Report 1461, MIT AI Lab, Massachusetts Institute of Technology, Cambridge, MA (1993)

[6] M. Turk and A. Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, Vol.3 (1), pp. 71–861.

[7] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, USA, (1999)

[8] Jacobs, R., Jordan, M., Nowlan, S., and Hinton, G., "Adaptive Mixtures of Local Experts", *Neural Comput3*, 1991, pp. 79–87

[9] R. Ebrahimpour, E. Kabir and M.R. Yousefi, "Face Detection Using Mixture of MLP Experts", *Neural Processing Letters*, 2007, pp. 69-82.

[10] N. Taheri M., A. Hajiany, R. Ebrahimpour G. Sepidnam, "A Modified Mixture of MLP Experts for face recognition", To appear in Proceedings of IPCV 2008, Las Vegas, Nevada, USA.

[11] Zhang M, Flucher J. Face recognition using artificial neural network Group-based Adaptive Tolerance (GAT) trees. IEEE Transactions on Neural Networks 1996; 7(3):555–567

[12] R. Ebrahimpour, E. kabir, M. Yousefi, "Teacher-directed learning in view-independent face recognition with mixture of experts using single-view eigenspaces", *Franklin Institute,* 345, 2008, pp.87–101.

[13] MathWorks, Inc., 2007. MATLAB. 3 Apple Hill Drive, Natick, MA, USA.

[14] Turk, M. and Pentland, A.: Eigenfaces for Recognition. J. Cognitive Neurosci. **3** 1 (1991) 71-86

Dependability Comparison of Explicit and Numerical GPC Algorithms

Piotr Gawkowski*, Maciej Ławryńczuk†, Piotr Marusak†, Piotr Tatjewski† and Janusz Sosnowski*
*Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland
{P.Gawkowski, J.Sosnowski}@ii.pw.edu.pl
†Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland
{M.Lawrynczuk, P.Marusak, P. Tatjewski}@ia.pw.edu.pl

***Abstract*- The paper presents the dependability comparison of software implementation of the explicit and numerical Generalized Predictive Control (GPC) Model Predictive Control (MPC) algorithms. The investigated GPC algorithms are implemented for a control system of a multivariable chemical reactor – a process with strong cross-couplings. The fault sensitivity of the proposed implementations is verified in experiments with a software implemented fault injector. Experimental methodology of disturbing the software implementation of control algorithms is presented. The influence of faults on the quality of the controller performance is analysed. Implementations of the considered control algorithms are also compared in terms of their profiles (e.g. resource usage, code size and efficiency).**

I. INTRODUCTION

The fault influence on the control performance is thoroughly discussed in the literature, see [12] and references therein. Typically, considered fault models are related to some kinds of malfunctions within the controlled system (e.g. actuators and sensor failures, faults in the plant itself). The influence of faults affecting software implementations of control algorithms is not a frequent subject of research. Recent studies show that the probability of fault occurrence in the microprocessor devices increases in new technologies as the size of transistor gates shrinks [18]. These faults may result in logical errors at the application level and finally in application failures. Therefore, faults that appear during system operation may be critical [10], particularly in case of control systems where faults affecting the code with implementation of the control algorithm can negatively influence economic efficiency or even safety of the process [3, 4, 5]. Thus, it is extremely important to make the algorithms as robust to such faults as possible.

Model Predictive Control (MPC) algorithms are widely used in the industry thanks to their numerous advantages [13-17, 20]. More specifically, they can offer better performance than the one offered by classical control algorithms, especially for processes with multiple inputs and multiple outputs (MIMO) and with significant time delay. Furthermore, when necessary, MPC algorithms can take into account constraints imposed on manipulated and controlled variables.

This paper compares the dependability of software implementation of the Generalized Predictive Control (GPC) algorithms [2] applied to a multidimensional chemical reactor [1]. Two versions of the GPC scheme are considered: the explicit and the numerical ones. In the first case the control policy is explicitly calculated. If necessary, the constraints can be taken into account by appropriate clipping of manipulated variable values. In the numerical GPC algorithm, constraints are taken into account in the algorithm. Such an approach needs solving on-line the quadratic programming problem.

In the research carried out the software implemented fault injector adapted to reactive applications is used [6, 9, 10, 11, 19]. This paper continues previous research of authors during which preliminary dependability studies of explicit MPC algorithms were conducted and fault hardening approaches proposed [3, 4, 5].

In the next section both the explicit and the numerical GPC control algorithms are described. Section 3 presents software fault injector and experiment set-up. Then, section 4 describes considered software implementations. Some software improvements towards fault hardening are also presented. The dependability of the analysed implementations based on experimental results are presented in section 5. The paper is shortly summarized in the last section.

II. THE GPC ALGORITHM

In the GPC algorithm [14, 17, 20] a dynamic model of the process is used in order to predict its future behaviour and to calculate the optimal control policy. More specifically, at each consecutive sampling instant k a set of future control increments

$$\Delta\boldsymbol{u}(k) = [\Delta u(k\,|\,k) \quad \Delta u(k+1\,|\,k) \quad \dots \quad \Delta u(k+N_u-1\,|\,k)]^T \tag{1}$$

is calculated assuming that $\Delta u(k+p\,|\,k)=0$ for $p \geq N_u$, where N_u is the control horizon. In the following part of the article it is assumed that the process has n_u inputs and n_y outputs ($u \in \Re^{n_u}$, $y \in \Re^{n_y}$). The future control policy $\Delta\boldsymbol{u}(k)$ is calculated in such a way that future predicted control errors (i.e. differences between the desired trajectory and predicted values of the output) are minimised over the prediction horizon N. For optimisation the following quadratic cost function is usually used

$$J(k) = \left\| \boldsymbol{y}^{ref}(k) - \hat{\boldsymbol{y}}(k) \right\|_{\boldsymbol{M}}^2 + \left\| \Delta\boldsymbol{u}(k) \right\|_{\boldsymbol{\Lambda}}^2 \tag{2}$$

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_76,

where the reference trajectory $\boldsymbol{y}^{ref}(k)=\left[y^{ref}(k+1\,|\,k) \quad \dots \quad y^{ref}(k+N\,|\,k)\right]^T$ and the prediction $\hat{\boldsymbol{y}}(k)=\left[\hat{y}(k+1\,|\,k) \quad \dots \quad \hat{y}(k+N\,|\,k)\right]^T$ are vectors of length n_yN, $\boldsymbol{M}$ and $\boldsymbol{\Lambda}$ are weighting matrices of dimensionality $n_yN\times n_yN$ and $n_uN_u\times n_uN_u$, respectively. Usually, the reference trajectory is assumed constant over the prediction horizon, i.e. $\boldsymbol{y}^{ref}(k)=\left[y^{ref}(k) \quad \dots \quad y^{ref}(k)\right]^T$.

Although in the GPC algorithm a number of future control increments (1) over the control horizon are calculated, only the first n_u elements of the vector $\Delta\boldsymbol{u}(k)$ are actually applied to the process, i.e.

$$u(k)=\Delta u(k\,|\,k)+u(k-1) \tag{3}$$

At the next sampling instant, k+1, the prediction is shifted one step forward and the whole procedure aiming at optimising the future control policy is repeated.

In the GPC algorithm, predictions $\hat{y}(k+p\,|\,k)$ over the prediction horizon (p=1,...,N) are calculated using a linear dynamic model of the process

$$\boldsymbol{A}(z^{-1})y(k)=\boldsymbol{B}(z^{-1})u(k) \tag{4}$$

where

$$\boldsymbol{A}(z^{-1})=\begin{bmatrix} 1+a_1^1z^{-1}+\ldots+a_{nA}^1z^{-nA} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1+a_1^Mz^{-1}+\ldots+a_{nA}^Mz^{-nA} \end{bmatrix} \tag{5}$$

and

$$\boldsymbol{B}(z^{-1})=\begin{bmatrix} b_1^{1,1}z^{-1}+\ldots+b_{nB}^{1,1}z^{-nB} & \cdots & b_1^{1,N}z^{-1}+\ldots+b_{nB}^{1,N}z^{-nB} \\ \vdots & \ddots & \vdots \\ b_1^{M,1}z^{-1}+\ldots+b_{nB}^{M,1}z^{-nB} & \cdots & b_1^{M,N}z^{-1}+\ldots+b_{nB}^{M,N}z^{-nB} \end{bmatrix} \tag{6}$$

are matrices which comprise parameters a_i^m (m=1,...,n_y, i=1,...,n_A) and $b_i^{m,n}$ (m=1,...,n_y, n=1,...,n_u, i=1,...,n_B) of the model (integers n_A and n_B determine the order of the model).

Thanks to using the linear model of the process, it is possible to express the output prediction as the sum of a forced trajectory (which depends only on future control increments $\Delta\boldsymbol{u}(k)$) and a free trajectory (which depends only on the past)

$$\hat{\boldsymbol{y}}(k)=\boldsymbol{G}\Delta\boldsymbol{u}(k)+\boldsymbol{y}^0(k) \tag{7}$$

where $\boldsymbol{y}^0(k)=\left[y^0(k+1\,|\,k) \quad \dots \quad y^0(k+N\,|\,k)\right]^T$. The dynamic matrix $\boldsymbol{G}$ of dimensionality $n_yN\times n_uN_u$ is comprised of step-response coefficients of the model.

A. The Explicit GPC Algorithm

When the cost function (2) is optimised without any constraints, the future vector of optimal control increments is calculated explicitly

$$\Delta\boldsymbol{u}(k)=\boldsymbol{K}(\boldsymbol{y}^{ref}(k)-\boldsymbol{y}^0(k)) \tag{8}$$

where $\boldsymbol{K}=(\boldsymbol{G}^T\boldsymbol{G}+\Lambda)^{-1}\boldsymbol{G}$ is a matrix of dimensionality $n_uN_u\times n_yN$ which is calculated off-line.

$$u(k\,|\,k)=\boldsymbol{k}^e y^{ref}(k)+\sum_{j=1}^{n_B}\boldsymbol{k}_j^u\Delta u(k-j)+\sum_{j=0}^{n_A}\boldsymbol{k}_j^y y(k-j) \tag{9}$$

where $\boldsymbol{k}^e$, $\boldsymbol{k}_j^u$, $\boldsymbol{k}_j^y$ are matrices of appropriate dimensionality which are calculated off-line from the parameters of the model [20]. The obtained explicit control law is a linear feedback from the difference between the set-point trajectory and values of the manipulated variable increments calculated at previous sampling instants.

Although in the explicit GPC algorithm constraints are not taken into account, it is possible to clip the unconstrained solution calculated by means of (9).

B. The Numerical GPC Algorithm

In contrast to the explicit unconstrained GPC algorithm, when the constraints are imposed on process variables, at each sampling instant the following quadratic programming problem must be solved in order to obtain optimal future control increments $\Delta\boldsymbol{u}(k)$

$$\begin{gathered}\min_{\Delta\boldsymbol{u}(k)}\left\{J(k)=\left\|\boldsymbol{y}^{ref}(k)-\boldsymbol{G}\Delta\boldsymbol{u}(k)-\boldsymbol{y}^0(k)\right\|_{\boldsymbol{M}}^2+\left\|\Delta\boldsymbol{u}(k)\right\|_{\boldsymbol{\Lambda}}^2\right\} \\ -\Delta\boldsymbol{u}^{\max}\le\Delta\boldsymbol{u}(k)\le\Delta\boldsymbol{u}^{\max} \\ \boldsymbol{u}^{\min}\le\boldsymbol{u}(k)\le\boldsymbol{u}^{\max} \\ \boldsymbol{y}^{\min}\le\boldsymbol{G}\Delta\boldsymbol{u}(k)+\boldsymbol{y}^0(k)\le\boldsymbol{y}^{\max}\end{gathered} \tag{10}$$

Vectors $\Delta\boldsymbol{u}^{\max}$, $\boldsymbol{u}^{\min}$, $\boldsymbol{u}^{\max}$, $\boldsymbol{y}^{\min}$, $\boldsymbol{y}^{\max}$ denote constraints imposed on manipulated and controlled variables, respectively.

III. EXPERIMENT SET-UP

The concept of the Software Implemented Fault Injector (SWIFI) is based on the software emulation of a fault during the run-time of the application under test. In this research FITS fault injector is used [6, 11, 19]. It uses standard Win32 *Debugging API* to control the execution of the software application under tests. In the following the process controlled by the analysed control algorithms (GPC) is described, instrumentation (facilitating further analysis) of the tested application is introduced, then the fault insertion policy, and finally, applied result qualification are given.

A. MIMO Process Description

The process under consideration is a chemical reactor with two manipulated (inputs) and two controlled (outputs) variables shown in Fig. 1.

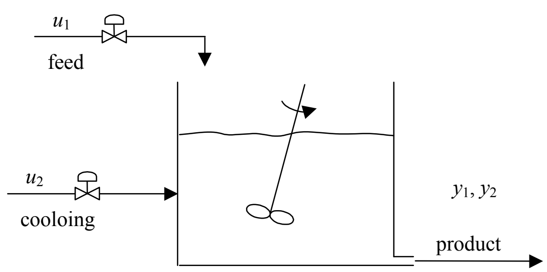


Fig. 1. The chemical reactor

It is described by the continuous-time transfer function model [1], (time constants in minutes):

$$\begin{bmatrix} Y_1(s) \\ Y_2(s) \end{bmatrix} = \begin{bmatrix} \frac{1}{1+0{,}7s} & \frac{5}{1+0{,}3s} \\ \frac{1}{1+0{,}5s} & \frac{2}{1+0{,}4s} \end{bmatrix} \begin{bmatrix} U_1(s) \\ U_2(s) \end{bmatrix} \tag{11}$$

where dimensionless controlled variables are: Y_1 – the concentration of the product, Y_2 – the temperature in the reactor, the dimensionless manipulated variables are: U_1 – the flow rate of the feed , U_2 – the flow rate of the cooling substance.

Using the sampling period T_p=0.03 min, one obtains the discrete-time dynamic model (4) with the following parameters

$$\boldsymbol{A}(z^{-1}) = \begin{bmatrix} 1-1.862885z^{-1}+0.866877z^{-2} & 0 \\ 0 & 1-1.869508z^{-1}+0.873715z^{-2} \end{bmatrix}$$

$$\boldsymbol{B}(z^{-1}) = \begin{bmatrix} 0.041951z^{-1}-0.037959z^{-2} & 0.475812z^{-1}-0.455851z^{-2} \\ 0.058235z^{-1}-0.054027z^{-2} & 0.44513z^{-1}-0.136097z^{-2} \end{bmatrix} \tag{12}$$

For the considered chemical reactor the explicit and the numerical GPC algorithms are designed. In both cases the prediction horizon is N=3, the control horizon N_u=2, the weighting matrices are: $\boldsymbol{M}=\boldsymbol{I}$, $\boldsymbol{\Lambda}=\lambda\boldsymbol{I}$, where λ=0,075. Magnitudes of manipulated variables are constrained: $u_1^{\min}=-2.5$, $u_1^{\max}=2.5$, $u_2^{\min}=-0.6$, $u_2^{\max}=0.6$, control increments are not constrained, values of manipulated variable are also not constrained.

The simulation horizon is 250 discrete time-steps. The simulation scenario is as follows:

- the process starts from a given set-point (y_1=0, y_2=0),
- at the sampling instant k=10 y_1^{ref} changes to 0.7 and y_2^{ref} to -0.5,
- at k=110 y_1^{ref} changes to – 0.7,
- at k=160 y_2^{ref} changes to 0.5.

Moreover, unmeasured disturbances are added to both outputs of the process. It is assumed that they change in a sinusoidal way (inverted on the second output variable) with the amplitude equal to 0.2 and the period of 50 discrete time instants, i.e.

$$y_i(k) = (-1)^i\, 0.2 \sin(0.04\pi \cdot i) \tag{13}$$

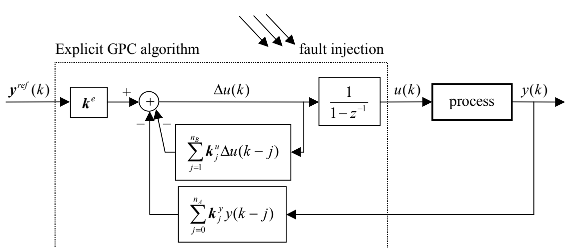


Fig. 2. The structure of the control system with the explicit GPC algorithm

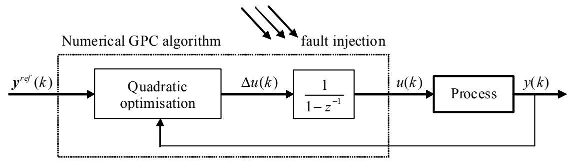


Fig. 3. The structure of the control system with the numerical GPC algorithm

B. *Code Instrumentation*

FITS disturbs directly the tested application only within so-called testing areas [6]. Testing areas limit the scope of disturbances only to the selected parts of the application. Here, the code of the controlled process model is added. Moreover, the controller code and data are moved to separate library module (DLL). In all experiments the same process modelling module is used together with different implementations of the controller module. The parts of the tested application disturbed during the experiments (dashed box) as well as process model (not disturbed) are marked in Fig. 2 and 3. The tested application is also instrumented to send some measures (e.g. related to internal variables, output signals) to the fault injector using user-defined messages (collected by FITS) [6].

C. *Fault Injection Policy*

FITS simulates faults by disturbing the running application. In this study the single bit-flip faults within CPU and FPU registers, application's data and machine instruction code are considered. In case of faults within instruction code the latching (fault is not recovered by the fault injector after activation) and non-latching (faulty instruction is recovered by fault injector after the first execution of disturbed instruction) strategies are used [9, 19]. Faults are injected pseudorandomly in time of program execution and in space (bit position within disturbed resource, distribution over application's memory) to mimic Single Event Upset (SEU) effects [9, 10, 19]. The time window for injecting faults is limited to the first half of simulation horizon (i.e. during first 125 sampling instants) to cover late error manifestation.

D. *Qualification of Experimental Results*

Control algorithms require more complex analysis of the controlled process behaviour than simple calculation-oriented ones [10, 19]. The standard factor ISE (Integrated Sum of Squared Errors) is used as a measure of result (y_1, y_2) correctness. The reference ISE value (obtained during non-faulty execution) is 9.4322 for explicit GPC and 9.4174 for numerical GPC implementation.

The whole experiment is conducted by FITS automatically. Analysis of fault effects requires detailed information upon the faults injected and the application behaviour. FITS provides details about every test (simulated fault injection). Hence, manual replay of the whole test execution can be done. Moreover, all the events and user messages occurring during the test are recorded. The tested application is instrumented to save its outputs (here simulation results, i.e. a set of control signals in subsequent sampling instants) into separate files for each test (file names are managed by FITS). This gives a possibility for post-experiment analysis of fault effects in the correlation with the injected fault and observed behaviour for each test. At the end of the experiment synthetic (aggregated) results for each fault location are given. In general, 4 classes of test results are distinguished:

- C: correct behaviour (ISE<15),
- INC: incorrect (unacceptable) behaviour (ISE≥15),
- S: test terminated by the system due to un-handled exception,
- T: timed-out test.

Additional U (user message signalisation) category denotes test results where some error detection mechanisms embedded into controller's code signal erroneous operation of the controller (with proper user message) and interrupt execution.

IV. Considered Software Implementations

A. *Classical Implementations*

Classical software implementations of the explicit (further denoted as *E*) and numerical GPC (denoted as *N*) algorithms for considered chemical reactor process are made in the C language and compiled with Microsoft Visual C++ 2005 SP1 with default optimization options (maximize speed - /O2). In particular, the /EHsc parameter is provided (enabling C++ exception handling without support for native Win32 Structured Exception Handling - SEH) even though there is no exception handling statements in the source code. It is worth to note the significant difference in the profiles of the binaries of these two GPC implementations (discussed in the sec. IV.C).

The explicit GPC algorithm is written as a static sequence of instructions (no loops are present). The matrices are not represented as tables – instead, sets of standalone variables (representing related elements of matrices) are introduced. Such implementation level optimization reduced the number of instructions needed to accomplish the analyzed task comparing to solution using tables representing matrices. The improvement is more than 20% in static code and 83% in the number of all executed instructions. Such optimization is very hard in case of numerical GPC algorithm, as constrained quadratic optimization procedure is usually written as universal one and uses vectors and matrices.

The distribution of mnemonics in the static code as well as in the executed instruction stream (dynamic) differ for explicit and numerical implementation. Floating point instructions constitute 95% of the static code image of the explicit GPC implementation as well as in the dynamic profile. In case of the numerical implementation these instructions are only in 23% of the static code and 27% in the dynamic profile. The majority of executed instructions are related to the following machine instructions (static; dynamic): *mov* (28%; 26%), *add* (9%; 11%), *fld* (7%; 8%), and *cmp* (7%; 8%).

Moreover, profiles of the CPU registers usage differ. The explicit implementation is relying mostly on floating point unit – general purpose CPU registers are rarely used (only EAX and ECX registers with the activity ratios at 2.5% and 0.7%, respectively). Contrary, the numerical version uses all general purpose CPU registers at the activity ratios higher than 60% [19].

B. *Fault Hardened Implementations*

Beside the classical software implementations some software improvements are introduced. The idea of the fault hardened versions (further denoted as HE and HN for explicit and numerical algorithms, respectively) is the same for explicit and numerical algorithms. It uses standard C++ exception handling statements (*try/catch(...)*) to capture exceptions occurring during the execution of the control algorithm code (the same as used in the classical implementation). In the exception-handling block the FPU unit is reset (with *_fpreset* function) and the control variable value is left unchanged from the previous control iteration. Compilation is made with SEH enabled (/EHa option) as well as with the support for FPU exceptions (/fp:except option).

Exception handling with such a forward recovery scheme significantly improves the probability of correct behavior but unfortunately the probability of the incorrect behavior rises as well [7–9, 11]. To mitigate this effect some error detection mechanisms might be included within the control algorithms (common for both implementations, explicit and numerical). As the considered fault model is limited to single faults, there is no reason to continue controller operation if the exception reoccurred. Thus, to prevent unrecoverable errors, the exception occurrence counter is introduced – only one recovery attempt is made otherwise controller signals failure (U-category behavior) and terminates.

In case of hardened explicit implementation (HE) additional error detectors are also introduced. Some infinite floating point values can be observed as a result of injected faults – they obviously are not signaled by the exception mechanisms. Hence, the *_finite* function is used to detect such value violations on the manipulated variables (programming exception is raised to engage previously described handling procedure). Moreover, the ISE value is collected within predefined time window (over 10 sampling instants) – if the ISE exceeds the threshold (here $ISE_{th}=10.0$) the exception is raised. In case of HE implementation the exception handling procedure also reinitializes all of the controller parameters.

In the case of numerical implementation the manipulated variables values should satisfy constraints (see sec. II and III.A) – in HN version, such violation rises an exception. Evaluating the dependability, it is found that in some cases the control algorithm may produce constant outputs – the controller obviously is locked and does not work (see Fig. 4). Hence, the control-locking detector is introduced. If there is no change on both controller outputs for more than 4 subsequent algorithm iterations and if at least one controller output is within the constrained area (not on boundaries), the difference between process outputs and given y^{ref} is checked. Exception is raised to signal erroneous controller behavior if this difference is higher then a predefined threshold.

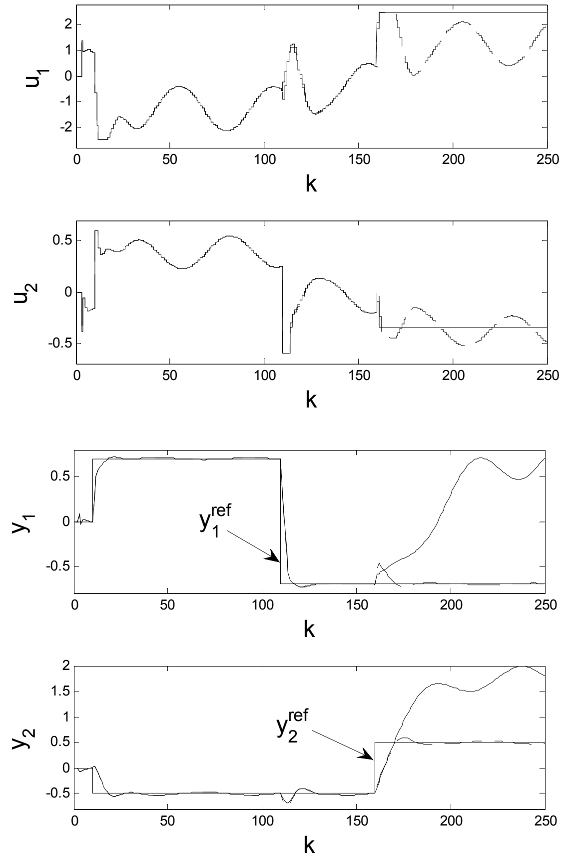


Fig. 4. Typical case with locked controller - fault-free trajectory (*dashed line*), erroneous trajectory (*solid line*), and the reference output trajectories (y_1^{ref} and y_2^{ref}) – fault injected at the sampling instant k=94 (*ISE*= 204.9).

C. Fault Hardening Overheads

Hardening code embedded into the classical implementation introduces some overheads in terms of time for computation (i.e. number of executed instructions within the controller) as well as in terms of larger static code size (in bytes and the number of instructions in the static code of the controller image). Nevertheless, there is also a significant difference of the code size and computation time between explicit and numerical GPC implementations. All these issues are covered in Table I.

TABLE I
STATIC AND DYNAMIC SODE SIZES AND CORRESPONDING OVERHEADS (IN BRACKETS).

	E	HE	N	HN
Static [bytes]	1050	1634 (56%)	12736	15327 (20%)
Static [instructions]	286	438 (53%)	3942	4531 (15%)
Dynamic [Instructions]	34522	54295 (57%)	2360990	2735855 (16%)

The relative overheads in the case of the explicit algorithm are much bigger than in the case of the numerical one. It is due to the fact that the code of numerical implementation is much more complicated comparing to the size of fault detection and tolerance routines. Whereas, the fault hardened implementation of the explicit algorithm is still 9, 10 and even 50 times more compact (in terms of the code size in bytes, the number of instructions and the number of dynamically executed instructions, respectively).

V. EXPERIMENTS

Faults in CPU and FPU registers, data area of the controllers, and the executed instruction stream (with latching and non-latching policy) are considered. For each fault location approximately 1000 disturbed executions are investigated (single fault injected in each application execution). Such faults within explicit GPC can even destabilize the controlled process [3, 5]. Here, we focus on comparison of explicit and numerical implementations as well as on the possible effectiveness of the proposed fault-tolerance mechanisms.

The summary of all 4 considered implementations is shown in fig. 5 and 6. It is clear that the explicit implementation is more fault robust than the numerical one. Nevertheless, the explicit one is less safe as the percentage of incorrect behaviour is slightly higher than observed in the numerical case. On the other hand, the explicit implementation is much more efficient if the code size is taken into account (see sec. IV.C). Thus, from the dependability perspective explicit GPC is better. However, hardened versions of both algorithms have similar level of fault robustness.

Fault detection and tolerance mechanisms applied proved to be very efficient, especially if faults are located in the resources that keep temporary values (e.g. CPU registers) or can be easily recovered (e.g. explicit GPC parameters in the data memory, non-latching faults within instruction stream). Only single incorrect test cases were observed for hardened versions (hard to notice in figures): 0.4% in the case of CPU registers faults in HE, 0.5% in the case of faults within data area in HN, and 2.3% in the case of non-latching faults in HE instructions.

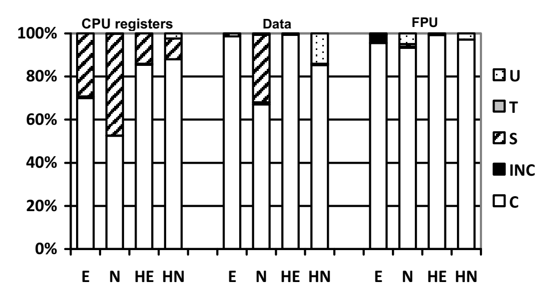


Fig. 5. Summary of experimental results (part 1).

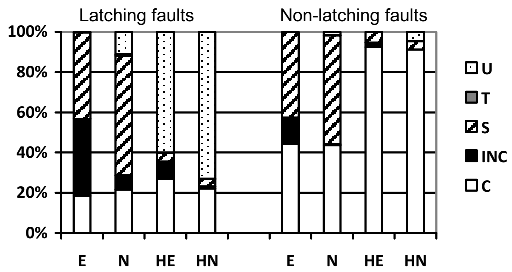


Fig. 6. Summary of experimental results (part 2) – faults within executed instruction stream.

The most fault sensitive resource is the controllers' code. Latching faults remain in the memory as any code recovery schema were not considered. That explains the difference between fault tolerance level of both hardened implementations with latching and non-latching policy. It is worth noting, that even if the level of correct behaviour for HE and HN implementations in case of latching faults within instructions is lower than 30% the hardened versions are much safer than classical ones. As error detection takes place (U-category), the reduction of incorrect behaviour (INC-category) is by the factor of 4.5 and 7 (in explicit and numerical implementations, respectively).

VI. CONCLUSIONS

The paper studies the dependability of software implementations of the explicit and numerical GPC algorithms applied to a chemical reactor. It is examined using software implemented fault injector. Results of the experiments clearly indicate that both formulations of the GPC algorithm are susceptible to hardware faults. It is particularly important in the case of industrial processes because faults may lead to undesirable behaviour of the process.

To achieve high degree of fault robustness several error detection mechanisms within controller code are embedded, each covering different erroneous situations. Those detectors are integrated with very simple but efficient recovery schema based on exception handling. The way to systematically eliminate incorrect cases requires detailed insight into the controller structure and analysis of process responses to identify the most typical controller deviations. That can be easily achieved with FITS fault injection system. Presented research proves that the software implemented fault detection and handling techniques proposed for both GPC algorithms can significantly improve their dependability.

ACKNOWLEDGMENT

This work was supported by a research grant from the Dean of the Faculty of Electronics and Information Technology, Warsaw University of Technology.

REFERENCES

[1] E. F. Camacho, C. Bordons, *Model Predictive Control*. Springer, 1999.

[2] D. W. Clarke, C. Mohtadi, P. S. Tuffs, "Generalized predictive control – I. The basic algorithm – I. The basic algorithm," *Automatica*, 1987, vol. 23, pp. 137-148, 1987.

[3] P. Gawkowski, M. Ławry czuk, P. Marusak, J. Sosnowski, P. Tatjewski, *Software Implementation of Explicit DMC Algorithm with Improved Dependability*, (T. Sobh, K. Elleithy, A. Mahmood, M. A. Karim, Eds.). Springer. London, 2008, pp. 214-219 (International Joint Conferences on Computer, Information and Systems Sciences and Engineering CISSE 2007).

[4] P. Gawkowski, M. Ławry czuk, P. Marusak, J. Sosnowski, P. Tatjewski, "Dependability of the explicit DMC algorithm for a rectification process," Proceedings of the International Multiconference on Computer Science and Information Technology, Wisła, Poland, pp. 779-788, 2007.

[5] P. Gawkowski, M. Ławry czuk, P. Marusak, J. Sosnowski, P. Tatjewski, "Dependability of explicit DMC and GPC algorithms," Proceedings of the International Multiconference on Computer Science and Information Technology, Wisła, Poland, pp. 903-912, 2007.

[6] P. Gawkowski, J. Sosnowski, "Experiences with software implemented fault injection," Proceeding of the International Conference on Architecture of Computing Systems, VDE Verlag GMBH, pp. 73-80, 2007.

[7] P. Gawkowski, J. Sosnowski, "Software implemented fault detection and fault tolerance mechanisms – part I: Concepts and algorithms," *Kwartalnik Elektroniki i Telekomunikacji*, vol. 51, pp. 291-303, 2005.

[8] P. Gawkowski, J. Sosnowski, "Software implemented fault detection and fault tolerance mechanisms – part II: Experimental evaluation of error coverage," *Kwartalnik Elektroniki i Telekomunikacji*, vol. 51, pp.495-508, 2005.

[9] P. Gawkowski, J. Sosnowski, B. Radko, "Analyzing the effectiveness of fault hardening procedures," Proceedings of the 11th IEEE Int. On-Line Testing Symposium, pp. 14-19, 2005.

[10] P. Gawkowski, J. Sosnowski, "Dependability evaluation with fault injection experiments," *IEICE Transactions on Information & System*, vol. E86-D, pp. 2642-2649, 2003.

[11] P. Gawkowski, J. Sosnowski, "Experimental Validation Of Fault Detection And Fault Tolerance Mechanisms," Proceedings of the 7th IEEE International Workshop on High Level Design Validation And Test, Cannes, France, 2002.

[12] J. Korbicz, J. M. Ko cielny, Z. Kowalczuk, W. Cholewa, *Fault Diagnosis Models, Artificial Intelligence, Applications*, Springer, London, 2004.

[13] M. Ławry czuk, P. Marusak, P. Tatjewski, "Multilayer and integrated structures for predictive control and economic optimisation," Proceeding of the IFAC/IFORS/IMACS/IFIP Symposium on Large Scale Systems: Theory and Applications, Gda sk, Poland, CD-ROM, paper 60, 2007.

[14] J. M. Maciejowski, *Predictive Control with Constraints*. Prentice Hall. Harlow, 2002.

[15] M. Morari, J. H. Lee, "Model predictive control: past, present and future," *Computers and Chemical Engineering*, vol. 23, pp. 667-682, 1999.

[16] S. J. Qin, T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Engineering Practice*, vol. 11, pp. 733-764, 2003.

[17] J. A. Rossiter, *Model-Based Predictive Control*. CRC Press, Boca Raton. 2003.

[18] N. Seifert, X. Zhu, L. W. Massengill, "Impact of Scaling on Soft-Error Rates in Commercial Microprocessors", IEEE Trans. On Nuclear Science, Vol. 49, No. 6, December 2002, 3100-3106.

[19] J. Sosnowski, P. Gawkowski, A. Lesiak, "Fault injection stress strategies in dependability analysis," *Control and Cybernetics*, vol. 33, pp. 679-699, 2005.

[20] P. Tatjewski, *Advanced Control of Industrial Processes, Structures and Algorithms*. Springer. London, 2007.

Efficient Predictive Control Algorithms Based on Soft Computing Approaches: Application to Glucose Concentration Stabilization

Maciej Ławryńczuk, Piotr M. Marusak, Piotr Tatjewski
†Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland
{M.Lawrynczuk, P.Marusak, P. Tatjewski}@ia.pw.edu.pl

***Abstract*- The paper presents computationally efficient predictive control approaches to stabilization of the blood glucose concentration in patients suffering from diabetes type 1. Presented algorithms use neural and fuzzy models and quadratic programming. These algorithms offer much better control performance than the algorithms based on linear models. Moreover, their closed–loop accuracy is similar to that obtained in predictive control algorithms with full nonlinear optimisation repeated on-line. Though simple, such algorithms offer advantages resulting from its prediction capabilities.**

I. INTRODUCTION

Model Predictive Control (MPC) algorithms are widely used in the industry thanks to their numerous advantages [13, 17, 19]. More specifically, they can offer better performance than that offered by classical control algorithms, especially for processes with multiple inputs and multiple outputs (MIMO) and with significant time delay. Furthermore, when necessary, MPC algorithms can take into account constraints imposed on manipulated and controlled variables.

When the considered process is nonlinear, it is justified to use its nonlinear model in MPC algorithm. In general, two approaches to nonlinear MPC can be distinguished. In the first one the full nonlinear model is used in MPC without any simplifications [13, 19]. Although in theory such an approach may seem to be likely to offer very good control accuracy, it needs solving on-line a nonlinear optimisation problem. From the practical point of view MPC algorithm with full nonlinear optimisation have a few important disadvantages. First of all, nonlinear optimisation is very computationally demanding. Moreover, nonlinear optimisation is likely to terminate in shallow local minimum, there is no guarantee the algorithm finds the global optimal solution at each sampling instant.

In order to overcome drawbacks of the MPC with full nonlinear optimisation, various MPC schemes with on-line linearisation have been proposed. More specifically, these algorithms find on-line a local linearisation of the original nonlinear model [10, 12, 13, 19]. Thanks to it, it is possible to formulate the MPC optimisation problem as a quadratic programming one. Such a problem can be efficiently solved, its computational load is small. It is also very important that in case of quadratic programming problems the global optimum solution is always found, there is no risk of terminating in local minimum. Stabilisation of glucose concentration is an important challenge. The algorithms used must be reliable which is not the case for algorithms with nonlinear optimisation as presented, for example in [8, 21].

In the next section both the basic formulation of MPC algorithms is given. Sections 3 and 4 present descriptions of the neural MPC and fuzzy MPC, respectively. In section 4 control plant (diabetic patient) is described and simulation results obtained in control systems of this plant are given. The paper is concluded by the last section.

II. MPC ALGORITHMS

In MPC algorithms at each sampling instant a set of future control increments

$$\Delta\boldsymbol{u}(k) = \left[\Delta u(k\,|\,k)^T \quad \ldots \quad \Delta u(k+N_u-1\,|\,k)^T\right]^T \tag{1}$$

is found. It is done in such a way that a criterion should be minimised and constraints present in the control system fulfilled. These demands are expressed as the following optimisation problem [3, 11, 13, 17, 19]

$$\begin{aligned}
&\min_{\Delta\boldsymbol{u}(k)}\{J_{MPC}(k) = \sum_{p=1}^{N}\left\|y^{sp}(k+p\,|\,k) - y(k+p\,|\,k)\right\|^2_{\boldsymbol{M}_p} + \\
&\qquad\qquad + \sum_{p=0}^{N_u-1}\left\|\Delta u(k+p\,|\,k)\right\|^2_{\boldsymbol{\Lambda}_p}\}\\
&\text{subject to:}\\
&\quad u_{\min} \le u(k+p\,|\,k) \le u_{\max}, \qquad p=0,\ldots,N_u-1\\
&\quad -\Delta u_{\max} \le \Delta u(k+p\,|\,k) \le \Delta u_{\max}, \qquad p=0,\ldots,N_u-1\\
&\quad y_{\min} \le y(k+p\,|\,k) \le y_{\max}, \qquad p=1,\ldots,N
\end{aligned} \tag{2}$$

where N and N_u denote prediction and control horizons, respectively, $\boldsymbol{M}_p \geq 0$ and $\boldsymbol{\Lambda}_p > 0$ are diagonal weighting matrices of dimension $n_y \times n_y$ and $n_u \times n_u$. The set-point trajectory is usually assumed to be constant over the prediction horizon, i.e. $y^{sp}(k+p\,|\,k) = \hat{y}^{ss}$, $p=1,\ldots,N$, ($y(k+p\,|\,k)$ denotes the output prediction for the future sampling instant $k+p$, derived at current sampling instant k, etc.). If for the prediction a linear dynamic process model is used, the problem (2) is a standard quadratic optimisation problem. Its solution at each iteration yields the new control increment vector $\Delta\boldsymbol{u}(k)$.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_77,

If for the prediction a linear dynamic process model is used, the superposition principle can be applied [11, 17, 19] and the prediction can be decomposed into two components

$$\boldsymbol{y}(k)=\boldsymbol{G}\Delta\boldsymbol{u}(k)+\boldsymbol{y}^0(k)\,, \tag{3}$$

where

$$\boldsymbol{G}=\begin{bmatrix}\boldsymbol{S}_1 & \boldsymbol{0} & \dots & \boldsymbol{0}\\ \boldsymbol{S}_2 & \boldsymbol{S}_1 & \dots & \boldsymbol{0}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{S}_N & \boldsymbol{S}_{N-1} & \dots & \boldsymbol{S}_{N-N_u+1}\end{bmatrix} \tag{4}$$

is the dynamic matrix of the dimension $n_yN\times n_uN_u$, composed of step response coefficients. The component $\boldsymbol{G}\Delta\boldsymbol{u}(k)$ is the forced response depending only on the future input increments, $\boldsymbol{y}^0(k)$ is the free response vector of the length n_yN depending only on the past

$$\boldsymbol{y}^0(k)=\left[y^0(k+1\,|\,k)^T \quad \dots \quad y^0(k+N\,|\,k)^T\right]^T \tag{5}$$

If output constraints are present in the optimisation problem (2), the infeasibility problems may occur. Thus, the output constraints are usually softened by the use of slack variables [15]. Assuming the quadratic penalty of the constraint violations, the MPC optimisation problem (2) with a linear process model is then as follows

$$\begin{aligned}&\min_{\Delta\boldsymbol{u}(k),\ \boldsymbol{\varepsilon}_{\min}\ \boldsymbol{\varepsilon}_{\max}}\{J_{MPC}(k)=\left\|\boldsymbol{y}^{ss}-\boldsymbol{G}\Delta\boldsymbol{u}(k)-\boldsymbol{y}^0(k)\right\|^2_{\boldsymbol{M}}+\\ &\qquad+\left\|\Delta u(k)\right\|^2_{\boldsymbol{\Lambda}}+\rho_{\min}\left\|\boldsymbol{\varepsilon}_{\min}\right\|^2+\rho_{\max}\left\|\boldsymbol{\varepsilon}_{\max}\right\|^2\}\\ &\text{subject to:}\\ &\boldsymbol{u}_{\min}\le\boldsymbol{J}\Delta\boldsymbol{u}(k)+\boldsymbol{u}^{k-1}\le\boldsymbol{u}_{\max}\\ &-\Delta\boldsymbol{u}_{\max}\le\Delta\boldsymbol{u}(k)\le\Delta\boldsymbol{u}_{\max}\\ &\boldsymbol{y}_{\min}-\boldsymbol{\varepsilon}_{\min}\le\boldsymbol{G}\Delta\boldsymbol{u}(k)+\boldsymbol{y}^0(k)\le\boldsymbol{y}_{\max}+\boldsymbol{\varepsilon}_{\max}\\ &\boldsymbol{\varepsilon}_{\min}\ge 0,\boldsymbol{\varepsilon}_{\max}\ge 0\end{aligned} \tag{6}$$

where $\rho_{\min}$, $\rho_{\max}$ are positive weights,

$$\begin{aligned}&\boldsymbol{u}_{\min}=\left[u^T_{\min}\quad\dots\quad u^T_{\min}\right]^T,\quad \boldsymbol{u}_{\max}=\left[u^T_{\max}\quad\dots\quad u^T_{\max}\right]^T\\ &\boldsymbol{u}^{k-1}=\left[u(k-1)^T\quad\dots\quad u(k-1)^T\right]^T\\ &\Delta\boldsymbol{u}_{\max}=\left[\Delta u^T_{\max}\quad\dots\quad\Delta u^T_{\max}\right]^T\end{aligned} \tag{7}$$

are vectors of the length n_uN_u and

$$\begin{aligned}&\boldsymbol{y}_{\min}=\left[y^T_{\min}\quad\dots\quad y^T_{\min}\right]^T,\quad \boldsymbol{y}_{\max}=\left[y^T_{\max}\quad\dots\quad y^T_{\max}\right]^T\\ &\boldsymbol{y}(k)=\left[y(k+1\,|\,k)^T\quad\dots\quad y(k+N\,|\,k)^T\right]^T\\ &\boldsymbol{\varepsilon}_{\min}=\left[\varepsilon_{\min}\quad\dots\quad\varepsilon_{\min}\right]^T,\ \boldsymbol{\varepsilon}_{\max}=\left[\varepsilon_{\max}\quad\dots\quad\varepsilon_{\max}\right]^T\end{aligned} \tag{8}$$

are vectors of the length n_yN and

$$\boldsymbol{J}=\begin{bmatrix}\boldsymbol{I}_{n_u\times n_u} & \boldsymbol{0}_{n_u\times n_u} & \boldsymbol{0}_{n_u\times n_u} & \dots & \boldsymbol{0}_{n_u\times n_u}\\ \boldsymbol{I}_{n_u\times n_u} & \boldsymbol{I}_{n_u\times n_u} & \boldsymbol{0}_{n_u\times n_u} & \dots & \boldsymbol{0}_{n_u\times n_u}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{I}_{n_u\times n_u} & \boldsymbol{I}_{n_u\times n_u} & \boldsymbol{I}_{n_u\times n_u} & \dots & \boldsymbol{I}_{n_u\times n_u}\end{bmatrix} \tag{9}$$

is a matrix of dimension $n_uN_u\times n_uN_u$. $\boldsymbol{M}$ and $\boldsymbol{\Lambda}$ are diagonal matrices of dimension $n_yN\times n_yN$ and $n_uN_u\times n_uN_u$ comprised of sub-matrices $\boldsymbol{M}_p$, $\boldsymbol{\Lambda}_p$, $\boldsymbol{y}^{ss}$ is a Kronecker tensor product of the vector $e^{ss}_y=\left[1\quad\dots\quad 1\right]^T$ of length n_u and y^{ss}, i.e.

$$\boldsymbol{y}^{ss}=e^{ss}_y\otimes y^{ss}=\left[\left(y^{ss}\right)^T\quad\dots\quad\left(y^{ss}\right)^T\right]^T \tag{10}$$

If the nonlinear model is used then (2) is the nonlinear optimisation problem which is in most cases time consuming and difficult to solve (see Fig. 1). Thus, algorithms using model linearisation were proposed. In the case of neural and fuzzy models obtaining of the linear approximation is relatively easy. It is described in details in the following part of the paper.

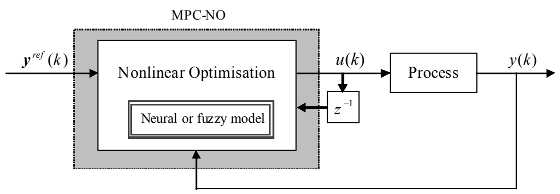


Fig. 1. The structure of the control system with the MPC algorithm with nonlinear optimisation (MPC–NO)

III. MPC ALGORITHMS BASED ON NEURAL AND FUZZY MODELS

The proposed MPC algorithms are the formulated in the following way:

1. The obtained linearised models (16) or (24) are used to generate the dynamic matrix and the elements of the control plant free response (5).
2. The free response and the dynamic matrix are used to formulate the quadratic optimisation problem (6).
3. The optimisation problem is solved and, using the obtained solution, the manipulated variable value is generated. Then the controller passes to the next iteration.

The graphical illustration of the algorithms is depicted in Fig. 2. In the case of neural and fuzzy models exploited by the algorithms under discussion its linear approximation is obtained in different ways. They are detailed in the next part of this section.

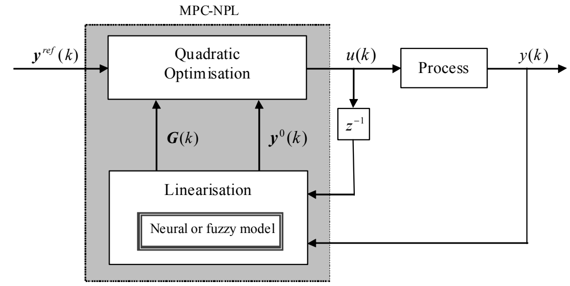


Fig. 2. The structure of the control system with the MPC algorithm with successive linearisation (MPC–SL)

A. Neural Network Model

The single-input single-output (SISO) neural model of the process is described by the following nonlinear discrete-time equation:

$$y(k) = f(u(k-\tau),\ldots,u(k-n_B), y(k-1),\ldots,y(k-n_A)) \tag{11}$$

where $f:\Re^{n_A+n_B-\tau+1}\rightarrow\Re$, $\tau\leq n_B$. A feedforward MLP neural network (Multi Layer Perceptron) with one hidden layer and a linear output [6] is used as the function *f* in (11). The structure of the neural model is depicted in Fig. 3. The output of the model can be expressed as:

$$y(k) = w_0^2 + \sum_{i=1}^{K} w_i^2 \varphi(z_i(k)) \tag{12}$$

where $z_i(k)$ are sums of inputs of the i^{th} hidden node, $\varphi:\Re\rightarrow\Re$ is the nonlinear transfer function (e.g. the hyperbolic tangent), K is the number of hidden nodes. Recalling input arguments of the general neural model (11) one has:

$$z_i(k) = w_{i,0}^1 + \sum_{j=1}^{I_u} w_{i,j}^1 u(k-\tau+1-j) + \sum_{j=1}^{n_A} w_{i,I_u+j}^1 y(k-j) \tag{13}$$

where $I_u = n_B - \tau + 1$. Weights of the network are denoted by $w_{i,j}^1$, $i=1,\ldots,K$, $j=0,\ldots,n_A+n_B-\tau+1$, and w_i^2, $i=0,\ldots,K$, for the first and the second layer, respectively. Combining equations (12) and (13) one obtains:

$$y(k) = w_0^2 + \sum_{i=1}^{K} w_i^2 \varphi(w_{i,0}^1 + \sum_{j=1}^{I_u} w_{i,j}^1 u(k-\tau+1-j) + \sum_{j=1}^{n_A} w_{i,I_u+j}^1 y(k-j)) \tag{14}$$

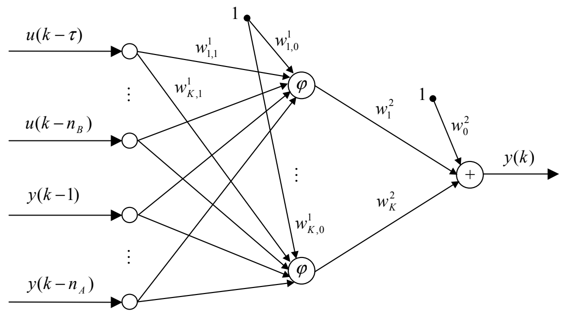


Fig. 3. The structure of the neural model

The linearisation point is defined as a vector comprised of past input, output and disturbance signal values corresponding to the arguments of the nonlinear model (11)

$$\bar{\boldsymbol{x}}(k) = \begin{bmatrix} x_1(k) \\ \vdots \\ x_{n_A+n_B-\tau+1}(k) \end{bmatrix} = \begin{bmatrix} u(k-\tau) \\ \vdots \\ u(k-n_B) \\ y(k-1) \\ \vdots \\ y(k-n_A) \end{bmatrix} \tag{15}$$

Using Taylor series expansion, the linearised model of the nonlinear model (11) derived at the current instant *k* is

$$y(k) = f(\bar{\boldsymbol{x}}(k)) + \sum_{l=1}^{n_B} b_l(\bar{\boldsymbol{x}}(k))(u(k-l) - x_l(k)) - \sum_{l=1}^{n_A} a_l(\bar{\boldsymbol{x}}(k))(y(k-l) - x_{n_A+\tau-1+l}(k)) \tag{16}$$

Coefficients of the linearised models are calculated from

$$a_l(\bar{\boldsymbol{x}}(k)) = -\frac{\partial f(\bar{\boldsymbol{x}}(k))}{\partial y(k-l)} \quad l=1,\ldots,n_A \tag{17}$$

$$b_l(\bar{\boldsymbol{x}}(k)) = \begin{cases} 0 & l=1,\ldots,\tau-1 \\ \dfrac{\partial f(\bar{\boldsymbol{x}}(k))}{\partial u(k-l)} & l=\tau,\ldots,n_B \end{cases} \tag{18}$$

Considering the neural model given by (14)

$$a_l(k) = -\sum_{i=1}^{K} w_i^2 \frac{d\varphi(z_i(\bar{\boldsymbol{x}}(k)))}{dz_i(\bar{\boldsymbol{x}}(k))} w_{i,I_u+l}^1 \quad l=1,\ldots,n_A \tag{19}$$

$$b_l(k) = \begin{cases} 0 & l=1,\ldots,\tau-1 \\ \displaystyle\sum_{i=1}^{K} w_i^2 \frac{d\varphi(z_i(\bar{\boldsymbol{x}}(k)))}{dz_i(\bar{\boldsymbol{x}}(k))} w_{i,l-\tau+1}^1 & l=\tau,\ldots,n_B \end{cases} \tag{20}$$

If hyperbolic tangent is used as the nonlinear transfer function φ in the hidden layer of the neural model, one has

$$\frac{d\varphi(z_i(\bar{\boldsymbol{x}}(k)))}{dz_i(\bar{\boldsymbol{x}}(k))} = 1 - \tanh^2(z_i(\bar{\boldsymbol{x}}(k))) \tag{21}$$

The linearisation point given by (15) is not influenced by the most recent output value $y(k)$, which is available. It may be crucial in the case of fast processes. Therefore, it is recommended to use

$$\bar{\boldsymbol{x}}(k)=\begin{bmatrix} x_1(k) \\ \vdots \\ x_{n_A+n_B-\tau+1}(k) \end{bmatrix}=\begin{bmatrix} u(k-\tau+1) \\ \vdots \\ u(k-n_B+1) \\ y(k) \\ \vdots \\ y(k-n_A+1) \end{bmatrix} \tag{22}$$

If τ=1 for linearisation purposes one may set $u(k\,|\,k):=u(k-1)$ or $u(k\,|\,k):=u(k\,|\,k-1)$.

B. Fuzzy Takagi–Sugeno Model

A method of synthesis of the Fuzzy MPC algorithm is detailed in [12], thus it will be now only outlined. Such an algorithm is very easy to design because it is based on DMC predictive algorithm [3, 5, 11, 17, 19] and fuzzy Takagi–Sugeno (TS) process model with local models in the form of step responses [12]. Such a model is relatively easy to obtain. It is sufficient to collect a few step responses of the control plant near a few operating points (three in the case under consideration). The membership functions can be chosen using expert knowledge, simulation experiments or fuzzy neural networks. The discussed model is thus described by the following rules:

$$\begin{aligned} &\text{If } y_k \text{ is } B_1^f \text{ and}\ldots\text{and } y_{k-n_p+1} \text{ is } B_{n_p}^f \text{ and}\ldots\text{and} \\ &\quad u_k \text{ is } C_1^f \text{ and}\ldots\text{and } u_{k-m_p+1} \text{ s } C_{m_p}^f \\ &\quad \text{then } \tilde{y}_{k+1}^f=\sum_{i=1}^{p_d-1} a_i^f\cdot\Delta u_{k-i}+a_{p_d}^f\cdot u_{k-p_d} \end{aligned} \tag{23}$$

where y_k is a value of the output variable at the k^{th} sampling instant, u_k is a value of the manipulated variable at the k^{th} sampling instant, $B_1^f,\ldots,B_{n_p}^f,C_1^f,\ldots,C_{m_p}^f$ are fuzzy sets, a_i^f (i=1,…,p_d) are the coefficients of step responses in the f^{th} local model describing influence of the input on the output, p_d is equal to the number of sampling instants after which the coefficients of the step responses can be assumed as settled, f = 1,…, l, l is number of rules.

Then, in the Fuzzy DMC (FDMC) algorithm, the TS model (23) is used in a smart way, at each iteration to obtain its linear approximation. For current sampling instant using: current values of process variables, the TS model (23) and fuzzy reasoning, the following model is obtained:

$$\tilde{y}_k=\sum_{i=1}^{p_d-1}\tilde{a}_i\cdot\Delta u_{k-i}+\tilde{a}_{p_d}\cdot u_{k-p_d} \tag{24}$$

where $\tilde{y}_k$ is the output of the control plant model at the k^{th} sampling instant, $\tilde{a}_i=\sum_{f=1}^{l}\tilde{w}_f\cdot a_i^f$, and $\tilde{w}_f$ are the normalised weights; see e.g. [14, 18]. In fact (24) is the step response control plant model valid for the current values of process variables.

IV. Simulation Results

The control plant under consideration is a patient with diabetes of type 1. It is a nonlinear plant. Its detailed model, often used by other researchers, one can find in [8]. The output of the plant y is glucose concentration, the manipulated variable u is administration of insulin, the main disturbance d=D_G is the amount of carbohydrates digested. Therefore it can be assessed and disturbance measurement can be used in the controller.

The controllers were designed in such a way that output is maintained between y_{min}=4.5 mmol/l and y_{max}=9 mmol/l. The aim of the control system operation is to stabilise the glucose concentration near the level y_{sp}=6 mmol/l despite the disturbance occurrence.

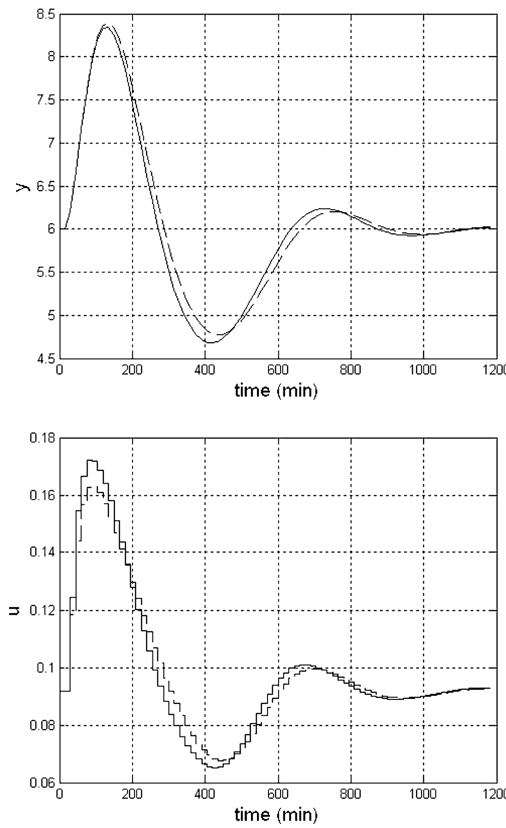


Fig. 4. Responses of the control system with MPC algorithms based on neural model to change of the disturbance; above – output variable, below – manipulated variable; MPC–NO – dashed line, MPC–SL – solid line

Fig. 4 shows the results obtained in the control systems with MPC algorithms based on the neural models. In the case of both algorithms the following parameters were assumed: sampling time T_p=15 minutes, prediction horizon N=20, control horizon N_u=5, λ=10. It is clearly visible that the algorithm with successive linearisation (MPC–SL) gives almost the same results as the MPC–NO one. It should be emphasised once again that the computational complexity of the MPC–SL algorithm is much smaller than of the MPC–NO one thanks to avoidance of a nonlinear optimisation.

The MPC–SL algorithm was then tuned in order to obtain bigger minimal value (the one obtained in the previous experiment was very close to the lower output constraint). The change of the prediction horizon to N=40 resulted in better trajectories as it can be easily seen in Fig. 5.

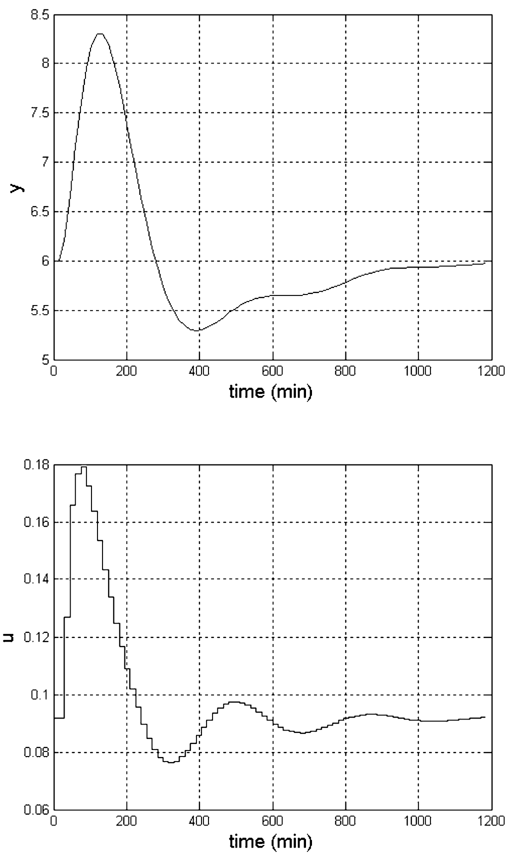


Fig. 5. Responses of the control system with MPC–SL algorithm based on neural model to change of the disturbance; above – output variable, below – manipulated variable; N=40

The Takagi–Sugeno fuzzy model used in the FDMC algorithm is the combination of three normalised step responses obtained from environs of the following operating points: u=0.092, y=6; u=0.077, y=9.95; u=0.112, y=4.043; these responses are shown in Fig. 6. The assumed membership functions are shown in Fig. 7.

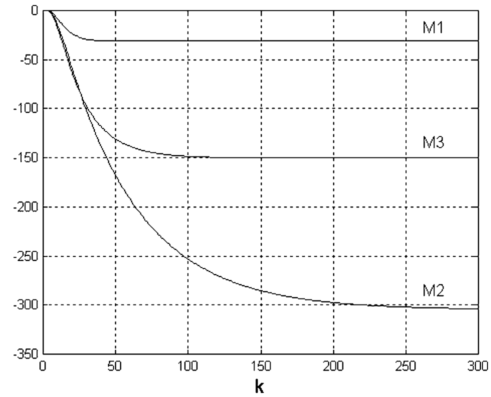


Fig. 6. Normalised step responses used in the TS fuzzy model

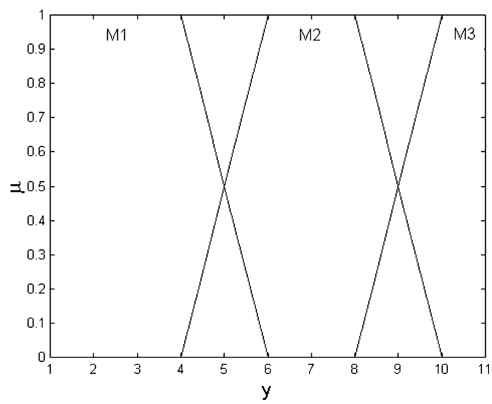


Fig. 7. Membership functions of the TS fuzzy model used by the FDMC algorithm

The sampling time of the FDMC algorithm T_p=15 minutes was assumed. The tuning parameters of the algorithm were assumed as follows: dynamics horizon p_d=300, prediction horizon N=20, control horizon N_u=5, λ=10.

The responses obtained in the control system with fuzzy MPC algorithm to changes of the disturbance are presented in Fig. 8. The controller maintains the output within the assumed range of values (dashed lines). It is good to notice that the obtained responses are pretty similar to those obtained with the algorithm based on the neural model (Fig. 5). The maximal values of the output are the same as well as the character of the responses. The differences which can be observed result from the different models used (in the case of fuzzy model the linear model in the form of step response is exploited whereas in the case of neural model – it is the difference equation).

In the another experiment the disturbance measurement was added to the controller. The obtained result (solid lines in Fig. 8) is much better. The maximal value of the output decreased from about 8.2 mmol/l to about 7.3 mmol/l. Moreover, the output stabilises faster and the minimal value of the output is around 6 mmol/l whereas in the previous case (without disturbance measurement) it was equal about 5.7 mmol/l.

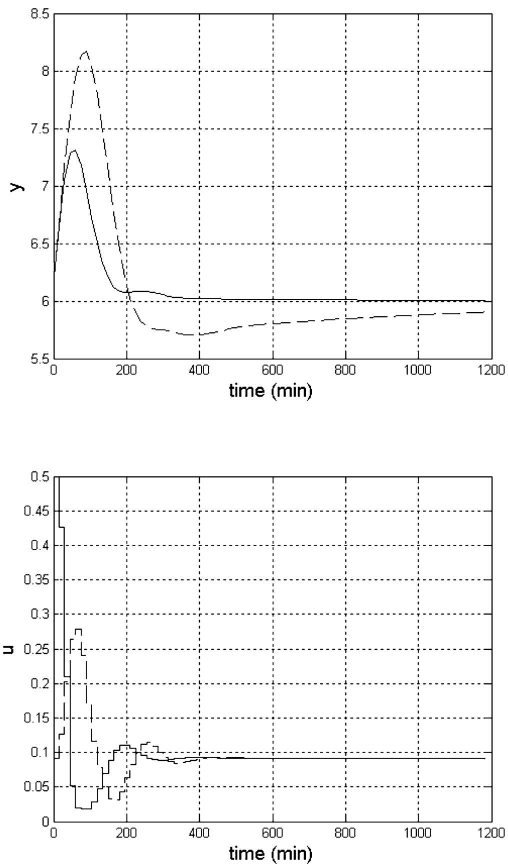


Fig. 8. Responses of the control system with FDMC algorithm (MPC–SL algorithm based on the fuzzy model) to change of the disturbance; above – output variable, below – manipulated variable

V. Conclusions

The obtained results clearly demonstrate usefulness of the proposed algorithms based on neural network and fuzzy Takagi-Sugeno models. Considered algorithms can offer the quality of control comparable to that obtained with the algorithm based on nonlinear optimisation. At the same time they demand much less computations because they are formulated in such a way that only quadratic optimisation problem must be solved at each iteration. It is especially important in control systems in which the manipulated variable must be generated quickly and reliably.

Very good result gave the algorithm with disturbance measurement. In the case of predictive controllers it is especially easy to include it in the algorithm. It is sufficient to add properly formulated terms to the free response which is next used to formulate the algorithm in a standard way.

References

[1] M. S. Bazaraa , J. Sherali, K. Shetty, *Nonlinear programming: theory and algorithms*. Prentice Hall, 1999.

[2] M. Brdys, P. Tatjewski, *Iterative Algorithms for Multilayer Optimizing Control*, Imperial College Press, London, 2005.

[3] E. F. Camacho, C. Bordons, *Model Predictive Control*. Springer, 1999.

[4] D. W. Clarke, C. Mohtadi, P. S. Tuffs, "Generalized predictive control – I. The basic algorithm – I. The basic algorithm," *Automatica*, vol. 23, pp. 137-148, 1987.

[5] R. Cutler, B. Ramaker, "Dynamic matrix control – a computer control algorithm," *AIChE National Meeting*, Houston, 1979.

[6] S. Haykin, *Neural networks – a comprehensive foundation*. Prentice Hall. Englewood Cliffs, 1999.

[7] K. Hornik, M. Stinchcombe, H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, 1989.

[8] R. Hovorka, V. Canonico, L. J. Chassin, U. Haueter, M. Massi-Benedetti, M. Orsini Federici, T. R. Pieber, H. C. Schaller, L. Schaupp, T. Vering, M. E. Wilinska, "Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes," *Physiol Meas*, vol. 25, pp. 905–920, 2004.

[9] M. Ławry czuk, P. Marusak, P. Tatjewski, "Cooperation of Model Predictive Control with steady-state economic optimisation," *Control and Cybernetics*, vol. 37, 2008, pp.133–158.

[10] M. Ławry czuk, "A family of model predictive control algorithms with artificial neural networks," *International Journal of Applied Mathematics and Computer Science*, vol. 17, pp. 217–232, 2007.

[11] J. M. Maciejowski, *Predictive Control with Constraints*. Prentice Hall. Harlow, 2002.

[12] P. Marusak, "Advantages of an Easy to Design Fuzzy Predictive Algorithm: Application to a Nonlinear Chemical Reactor," *Proc. Int. Multiconference on Computer Science and Information Technology*, Wisla, Poland, 2007.

[13] M. Morari, J. H. Lee, "Model predictive control: past, present and future," *Computers and Chemical Engineering*, vol. 23, pp. 667-682, 1999.

[14] A. Piegat, *Fuzzy Modeling and Control*. Physica–Verlag, Berlin, 2001.

[15] S. J. Qin, T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Engineering Practice*, vol. 11, pp. 733-764, 2003.

[16] D. Radomski, M. Ławry czuk, P. Marusak, P. Tatjewski, "Modeling of glucose concentration dynamics for predictive control of insulin administration," submitted to *Biocybernetics and Biomedical Engineering*.

[17] J. A. Rossiter, *Model-Based Predictive Control*. CRC Press, Boca Raton. 2003.

[18] T. Takagi, M. Sugeno, "Fuzzy identification of systems and its application to modeling and control," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 15, pp. 116–132, 1985.

[19] P. Tatjewski, *Advanced Control of Industrial Processes, Structures and Algorithms*. Springer. London, 2007.

[20] P. Tatjewski, M. Ławry czuk, "Soft computing in model-based predictive control," *International Journal of Applied Mathematics and Computer Science*, vol. 16, pp. 101-120, 2006.

[21] Z. Trajanoski, " Neural Predictive Controller for Insulin Delivery Using the Subcutaneous Route," *IEEE Transactions On Biomedical Engineering*, vol. 45, pp. 1122-1134, 1998.

Eteacher – Interactive and Individualized Training of Software Applications in the Original user Interface

Peter Krumhauer
Ebereschenallee 11
14050 Berlin, Germany

Abstract – **Eteacher is a software training application that works on the original program user interface by 'demonstration and exercise'. The student is free to choose his pace of learning and exercise. If an exercise of a lesson is not done correctly, the program intervenes and, if requested, presents the right solution. Teaching speed and content of the lessons are adapted to the needs of the student.**

Training lessons can be established for any software systems. The design of a Training Unit is easy and fast using mostly standard spreadsheet operations in combination with a recorder for voice, key strokes and mouse movements.

1. INTRODUCTION

Currently, the state of the art for computerized training programs for software is relying mostly on short prerecorded video clips that are played as part of the training. These clips show examples of activities and then the user is asked to repeat the shown content in a simulated user interface. These activities are very limited, e.g. only the activation of one function by mouse click and they have to be performed exactly the same way it was demonstrated.

Here, we propose a different approach. The advantage of learning with a human tutor is the interactive nature of demonstration, exercise and feedback as well as the adaptation to a student's knowledge and pace. To replicate these concepts we developed Eteacher, an application that

- Works within the original user interface of the host program (the program to be taught),
- Follows the student's actions during an exercise and checks for correct completion,
- Intervenes during the exercise as soon as the student is not on a path of correct execution.

In order for this to work, Eteacher will have the following characteristics:

1. It has to be able to be tuned to work with a range of software solutions as broad as spreadsheets and CAD programs,
2. The setup of Training Units has to be simple, fast and easy to manage and maintain,
3. It has to be cognizant of the numerous variants and possibilities to carry out a given function,
4. It has to follow the student's mouse activities and keystrokes and determine at any step of the way whether the student is on the right track,
5. It has to have the ability to assess and track the student's abilities and gauge the dynamics of demonstration and exercise to optimize the learning progress. If too many errors are made, the student is prompted for a lower teaching speed and units with lower speed will be chosen.

To demonstrate these characteristics, two different software samples are presented:

- Microsoft Excel and
- Solid Edge, a 3D design program.
-

2. THE DESIGN PROCESS OF AN MS EXCEL UNIT EXAMPLE

Training Goal of the unit
The goal of the unit is demonstrated by its start screen (Fig. 1). It shows the raw data, in this case the monthly sales figures for 3 salesmen, as well the goal that has to be achieved as task for the unit:

1. Format text and sales (Bold, Red, 15p, Currency €),
2. Compute the quarterly sales by creating a formula that sums up the monthly sales,
3. Calculate a 10% bonus for the best salesman using an Excel formula,
4. Present sales and bonus in a bar chart.

Fig. 1 Raw Data and Training Goal

The design process of this unit in detail:

Establishing the Training Unit
To setup the above lesson, the following activities are required by specialized personnel

The job of the Host Designer
The Host Designer does the overall design for all units for the program to be trained on (the host). He defines the teaching content items and designs the units.

The job of the Tool Designer
Eteacher works with any host software system. That means that all kind of software systems can be taught with its help. However, some hosts need specific tools that have to be programmed. Excel for example requires the ability to turn a cur-

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_78, © Springer Science+Business Media B.V. 2010

sor position to the name of a cell (e.g., A1). The tool designer is the only software programmer in the development process.

Checking Results versus Checking Action Paths

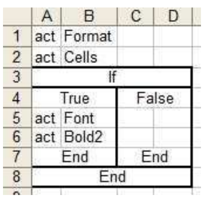
There are two fundamentally different ways of checking whether or not the student is carrying out an exercise correctly:

1. Tracking each activity (mouse or keyboard) and checking it against a predefined tree of possibilities, or
2. Giving the student freedom of manipulation and checking the results.

The former can be used for menu and toolbar button navigation, but to compose e.g. a formula for Excel is a complicated task with an infinite sequence of possible events of typing and deleting. These cases need result checking. Eteacher is capable of either method, as follows.

The Unit Graph: the navigation tree through a lesson

The basis of the unit is the unit graph – a spreadsheet structure that stores the actions of the unit:



	A	B	C	D
1	act	Format		
2	act	Cells		
3	If			
4	True		False	
5	act	Font		
6	act	Bold2		
7	End		End	
8	End			

Fig. 2 Actions with If- branching

Objects and Actions

Fig. 2 shows a sample of an action path of the unit with objects, actions and a branching:

- Objects are entities that can be activated by mouse click or key – in this case the button 'Format' on the toolbar
- Actions are activities that manipulate the objects – in this case 'act': activate – 'act Format' means 'activate the Format button'.
- The If branching decides the way of the action path dependent on a stored condition – if the condition is satisfied the True branch has to be followed, otherwise the False branch.

Each unit has a complex action tree with branches. The activities of the student are compared with this tree: leaving the tree means that something is carried out that will not solve the task of the unit and needs correction.

The job of the Function Designer

Structures for the unit may be of different complexity:

- Single objects – as Format (see above), or
- Complex trees like functions, Super-Objects or Super-Actions (shown later)

The job of designing a unit can be broken down to:

- All structures that might be used in another unit are done by the function designer,
- All structures that are specific for one unit are done by the unit designer.

Definition of Objects

The activation of objects by the student is checked by the program Analyzer. There are different types of objects, some of them are presented here:

The type Pixel Object monitors the change of color that happens when the cursor comes into the activation area of that object. When the student activates such an object the color of its pixels are compared with the previously stored activation color, the Verification Pixels.

The Verification Pixels can be defined by the function designer activating the button Define Object: The program presents a screen capture around the cursor position with visible single pixels, (see Fig. 3) and the Function Designer selects two representative pixels with the changed color, in this case from the button frame. In another accomplishment the Verification Pixels are found automatically if the unit designer moves the cursor to the object.

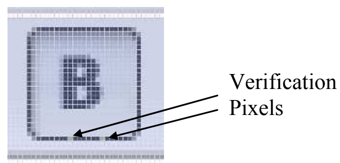


Fig. 3 Object Bold1 with 2 Verification Pixels in contrast color

The object is stored with name, type and Verification Pixels. In a similar way the Position- and Slide Object types are defined.

Other types of objects are Position Objects or Slide Objects (objects which can be selected using a slider), which are verified by checking if the cursor position was in the appropriate activation area.

Automatic Object Definition by the Crawler

The manual definition requires that the screen of the Function Designer and that of the student is identical, but this is given only in special cases where Eteacher runs in a controlled environment like a virtual machine or distant learning application.

If this is not the case the definition is done automatically by a crawler running on the computer of the student. For that to work, the crawler requires some predefined information about the topology of the host program user interface including menus and toolbar buttons. E.g., Excel has the standard menus File, Edit, View, etc. as well as a number of toolbars. The crawler finds the bars and buttons and by moving a cursor into the activation areas finds and stores automatically all necessary parameters of the object, like position, size, Verification Pixels, etc.

For objects invisible at the start screen but which appear following a given action path the crawler uses the menu structure of the host program that is prepared by the Function Designer. The object Bold2 for example (Bold1 is assigned to the Bold button) can be activated by the crawler following the action path Format/Cells/Font/Bold2.

Super-Actions and Super-Objects

Complex structures can be combined to and handled like a single object or a single action. This is demonstrated by a basic sample:

The marking of cell compound in Excel can be done either

- By dragging from the start cell to the end cell, or
- By activation of the start cell, holding Shift and activation of the end cell.

The representation for the action tree:

	A	B	C	D
1	act	A1		
2	Or			
3	Or1		Or2	
4	dr	B1	inky	Shift
5			act	B1
6	End		End	
7	End			

Fig. 4 Action tree for marking

This Super-Action is stored with its name, its action tree (Fig. 4) and the variable parameters (in this case B1 for the parameter A1 and D5 for parameter B1). The Super-Action 'Mark/A1/B1' and the structure in Fig. 4 are only different representations of the same action tree, but much easier to manage and maintain.

Similarly, Super-Objects can be defined based on a combination of objects, such as 'Bold2' in fig. 2, and variable parameters make them universally applicable. Super-actions and Super-Objects could again be part of a bigger structure (see below). The depth of such nesting is unlimited.

Functions

A number of features of any given program can be accessed using a variety of different ways through the Menu Tree, buttons or hotkeys. For example, there are three different ways to perform the task 'format a marked cell bold' in Excel:

- Activating the single object Bold1
- Following the action path Format/Cells/Font/Bold2
- Typing Ctrl+B.

The function Bold represents this using an action tree with an Or-branching between three alternatives. The object type is 'Function Object', which is a special Super-Object, representing all alternative solutions of the task. Hence, the action path 'Mark/A1/B1, act Bold' would ask the student to mark A1:B1 and then execute any of the above defined ways to format it Bold.

Control of Function Objects using the Menu Tree

Function Objects are not controlled like the other object types by following the actions of the student through the action tree in the unit graph but by following them through the Menu Tree: Parallel to the action tree, which is bound to the unit, is as second structure, the Menu Tree, which is fixed to the host program. It contains all objects and all branches of the Menu Tree of the program plus all other ways that do not use the menu buttons, like typing special key strings. The Menu Tree is generated in the same way as the action tree with actions, objects and branches.

Both structures are used to control the actions of the student: Generally, the student's actions are tracked by following the unit graph. But as soon as a function object is encountered, the control is passed to the Menu Tree.

This should be demonstrated on the following example. The actions of the student start with ST, the reactions of the Analyzer with AN.

The unit graph is initiated with the Super-Action Format A1/B1/Bold. The Analyzer expands it automatically to the action tree shown in Fig. 4 plus the following action act Italic.

ST	act A1
AN	ok!
ST	Hold Shift
AN	ok!
ST	act B1- A1:B1 are marked.
AN	ok!

The Analyzer now finds the next object Italic. In the object table its type is stored as Function Object and the Menu Tree takes now the control: there are 3 ways to the function Italic:

- The button Italic (in the Menu Tree the objects are not numbered)
- The path Format/Cells/Font/Italic
- The key combination Ctrl+I

ST	act Format (the student is apparently following the middle path of the Menu Tree in fig. 5.
AN	ok!
ST	act Cells
AN	ok!
ST	act Font
AN	ok!
ST	act Italic
AN	ok!
	Object Italic is done: now the unit graph takes over control again!

	A	B	C	D	E	F	G	H
1	Or							
2	Or1		Or2				Or3	
3	act	Italic1	act	Format			instr	Ctrl+I
4			act	Cells				
5			If					
6			True		False			
7					act	Font		
8			End		End			
9			End					
10			act	Italic2				
11			act	ok				
12	End		End				End	
13	End							

Fig. 5 Extract from the Menu Tree

Nesting of Function Objects

Function objects, like other super-entities, can be nested into others: The action path with the Super-Action 'Mark/A1/B1' and the function object 'act Bold' is condensed by the function designer to the Super-Action 'Format/A1/B1/Bold'. In an extended form the Super-Action would be 'Format/A1/B1/Bold/Red/15p' resulting in the task to format A1:B1 bold, red and 15p, all this with a free choice of the sequence of formatting.

The Super-Action "Formula/E4/E6/Quarter Sales"

In the example Training Unit, the cells E4 to E6 (see Fig. 1) are to be filled with the quarterly sales using a formula that sums up the monthly sales figures of the salesmen. Until now, the tracking of the student actions is done by tracking his way through the action tree with its And, Or and If branches.

The Super-Action Formula E4/E6/Quarter Sales checks:

- The input string to cell E4 generating the formula,
- This string against the string stored to the formula 'Quarter Sales'
- The drag from E4 to E6 to transfer the formula to the other salesmen cells.

Checking the input string against the standardized formula

The right formula for E4 is '=SUM(B4:D4)'. However, '= Sum (d4:b4)' with additional spaces, lower case letters and switched cell sequence would do just as well, while '=SUM(B4 : D4)' with spaces between the colon and the cell references is invalid.

To resolve these uncertainties, the input string is standardized to a default expression before the comparison. As the input can be done either by typing or by activation of objects (e. g. the button Sum or the cell B4) the formula is read from the input line using OCR.

The job of the Teacher

The demonstration of the unit, or lesson, is prepared by the teacher recording actions and spoken comments with the help of the program Recorder. The actions are detected and presented in the Lesson Graph (in a similar way as done by the Unit Generator):

Steps	Comments	Actions	Objects
1	Com1.1	act	A1
	Com1.2	dr	B1
	Com1.3	act	Bold1
	Com1.4	act	Fontcolor
	Com1.5	act	Red
	Com1.6	act	Fontsize
	Com1.7	act	15
	Com1.8	act	E3
	Com1.9	instr	Quarter Sales

…

Fig. 6 Lesson Graph (snapshot)

Editing the Lesson Graph

The compilation of objects and functions, their nesting, the voice-overs and the ability to do result checking requires the ability to optimize and refine a lesson. A lesson is organized into autonomous parts called steps that have an own stored start screen so that they might be resumed individually. A specialized tool called Lesson Editor allows the Teacher to edit the stream of events similarly to a movie editing application. A number of functions for browsing, navigating, stepwise debugging are provided.

Automatic transformation of fragments of the Action Paths of a lesson to the corresponding Super-Entities

Input for the unit designer is the lesson graph with the action path and the spoken comments of the Teacher. Fragments of the path, which represent just one way through the tree, are transformed automatically to the corresponding Super-Actions, Super-Objects and Functions to comprehensively assemble all possible ways through the tree.

To do this transformation all fragments of the path of the lesson graph are compared with all branches of all super-entities (Super-Objects, functions and Super-Actions) and if they are found to be identical the super-entities replace the fragments.

act	A1	act	A1		
dr	B1	Or			
		Or1		Or2	
		dr	B1	inky	Shift
				act	B1
		End		End	
		End			

Fig. 7 Comparison Lesson Graph and Super Action Mark/A1/B1

The comparison of the lesson graph (left in Fig. 6) and the Super-Action Mark/A1/B1 (right) shows that both actions of the lesson graph are identical with the left branch of the Super-Action. The identity of the cell references here is irrelevant as A1 and B1 are free parameters.

If in a lesson graph the path 'act cell1, dr cell2' is shown , or likewise 'act cell1, inky Shift, act cell2', then it is expanded automatically to the action tree with both possibilities of marking: The student, whose actions are tracked by the unit tree, is not bound to one given action path.

In the same way fragments with objects that represent function objects are transformed automatically to their function objects. E.g. the fragment 'act Bold1' is transformed to 'act Bold' by replacing the specific indicated way by the general way with all possibilities.

After further transformations (some more than one step) the Unit Graph looks like this:

Actions	Objects
Format	A1/B1/Bold/Red/15p
act	E3
instr	Quarter Sales
Formula	E4/E6/Quarter Sales
Format	B4/F6/Euro
act	F3
instr	Bonus
Format	B3/F3/Bold
Formula	F4/F6/Bonus
Chart	A3/F6

Fig. 8 Unit Graph expanded automatically from the Lesson Graph

The student is not obliged to follow this exact sequence, but could start with other actions, like the formatting of the sales 'instr Quarter Sales'. However, the sequence within the action chart is not free, at some point they all have to be performed. This freedom is realized by inserting And branches in the unit graph which provide the freedom to chose the sequence (not shown).

Assistance Required

Every time the student is uncertain what to do the button 'Assistance Required' can be invoked and the following choices are given:

- See all possible next steps
- Get an explanation
- Get an in-depth training

In-depth training must be part of the Unit Tree, in which case a link to such training is stored in the cell where the assistance is required, or an error occurred (see below).

In-depth Training

In-depth training is offered for complex situations, like for the training of formulas. In the sales figures example, a bonus of 10% is rolled out for the salesman with the highest quarter sales gets 10%. The right formula would be: =IF(E4=MAX(E$4:E$6); 10%*E12;0). This is a nested formula with the special formatting with '$'. Some students might understand this immediately, others need to get a special instruction before solving this task in the unit.

The job of the Unit Tester

The unit tester receives the unit from the unit designer, not knowing the host with the same abilities as the designated user. As there will be different units for users with different skill levels, there will be different testers.

During the test all of the actions are recorded as Test History. If the tester faces a problem it is recorded and fed back to the Unit Designer for improving the unit.

3. USE OF THE EXCEL UNIT BY THE STUDENT

Demonstration of a Lesson

At the beginning of the unit, the recorded lesson is performed demonstrating all actions of the teacher on the original screen of the host and voiced over with the comments of the teacher.

Exercising

In the next step the student can exercise himself what's just been demonstrated. The oversight over his actions is done by comparing them with the Action Tree of the unit. To allow this, the condensed unit graph is expanded automatically to the Action Tree with all branches that had been condensed before.

While in exercise mode, the student is only allowed to follow the tree. As soon as he does something outside the actual branch, an error message is generated and called out by context-sensitive comments (see below for detail).

If the student does not want to wait with the exercise until the end of the demonstration he can tell the program to start it right away. E.g. he could decide to repeat every action of the teacher immediately, possibly in a split screen mode.

Carry out of a similar task without prior demonstration

A student can also choose to work on a similar task after a completed exercise without the preceding demonstration of the teacher. This is controlled by its stored action tree as well.

4. THE DESIGN OF THE SOLID EDGE UNIT

Solid Egde is a CAD software that is significantly more complex that Excel. Still, Eteacher is able to demonstrate and oversee lessons in exactly the same simple fashion. Let's assume that the task of the Solid Edge Unit is to design a container with handles:

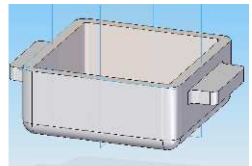


Fig. 9 Container

Condensed Unit graph

We start this unit in the end phase of the Unit Designer. As discussed before, the result of this phase is a list of condensed actions. The start of the unit is shown as in this snapshot:

Actions	Objects
act	Protrusion/Coincident/Xy
act	Cuboid/120/120/80/-/Sym/Sym
round	10/frlf/frri/frdw/bclf/bcri/bcdw/lfdw/ridw
thin	8/up
act	Protrusion/Parallel/Xy/-10/10

Fig. 9 Unit Graph Container (snapshot)

The graph shows actions and Super-Objects with rather complex tasks:

act Cuboid/120/120/80/-/Sym/Sym:

Design cuboid with a length of 120, being symmetrical on the x and y axes to the sketch center and z-borders with length of 80 without symmetry.

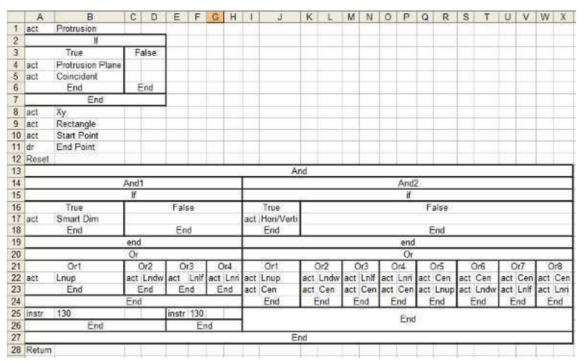


Fig. 10 Start of the Action Tree for the Container

Control of the Actions of the Student in the Repetition Phase
Expanded Unit Graph as basis of Student Control

To allow the tracking of the activities of the student, the condensed actions in the unit graph are now automatically expanded to the Action Tree with all branches (Fig. 10 shows a snapshot from the start).

The action tree above starting in A13 shows the variety of ways to define the dimensions and symmetries of the rectangle that forms the basis of the cuboid. We show this in the following real-life example. At the start, the action tree looks like this:

	A	B	C	D
1	act	Protrusion		
2	If			
3	True		False	
4	act	Protrusion Plane		
5	act	Coincident		
6	End		End	
7	End			
8	act	Xy		
9	act	Rectangle		
10	act	Start Point		
11	dr	End Point		
12	Reset			

Fig. 11 Action Tree Protrusion

The actions of the student are recorded as ST and the reactions of the Analyzer Program as AN:

AN	The only possible action for the student at the beginning of the repetition is 'act Protrusion'. The type of the object Protrusion is Pixel Object.
ST	act Protrusion – the Protrusion button changes its color
AN	The Analyzer compares stored colors of the Verification Pixels with those of the students screen. If they are identical the comment from the Analyzer will be 'ok'. Cells A1:B1 are marked as done

The next entity is the IF cell. Its stored condition is 'If (act(Coincident Plane)=0)' as shown above. The Analyzer checks the content in the object table and finds the object Coincident Plane not active: The True branch has to be followed and 'act Protrusion Plane' is marked as next possible action (see Fig.11)

ST	act Xy
AN	Error message and comment: 'Wrong step: Do you want to retry / F1 or get help / F2?'
ST	F2
AN	'Do you want me to perform the correct action / F1?' 'Do you want to get an explanation / F2?'
ST	F1
AN	Demonstrates 'Last Action' showing the cursor moving to the window 'Protrusion Plane' and how the window pops up
	'ok / F1 or do you need further help / F2?'
ST	F1
	act Protrusion Plane
AN	Ok
	A7 (End) is marked as last finished cell.
	The next actions (A8…A11) are correct.

Reset of Screen

The tracking of the student's actions is based on the expectation that the position of the objects on the screen of the student is the same as that of the teacher. In the case of designing a rectangle as basis for the cuboid (see Fig. 9) this is not the case as such sketches are done with the right topology but start and end point of teacher and student are different and therefore all lines will have different positions and the verification colors will be different.

To avoid such problems a Reset is inserted: The program issues the comment 'Please let me reset your screen to be equivalent to mine' and resets the screen. The rectangle of the teacher is shown and if the student now activates a line (which turns red) the colors of the Verification Pixels of student and teacher are identical. Objects like Return (which create new screens) do this automatically without the prompt.

The rest of the lesson is followed in the same way as described before: Next possible actions are found, marked and compared with that of the student.

5. CONCLUSION

Software training using the Eteacher is like learning with a private tutor:

The application demonstrates an activity accompanied by spoken comments. Then the student can repeat this activity on his own using his own way of doing it, he is not bound to the exact same sequence and the specific solutions of the teacher.

The lectures are adapted to the knowledge and the learning skills of the student comparing the knowledge profiles of the student and that of the units.

If the student feels uncertain what to do next or if he makes a mistake, the teacher offers assistance, which again is adapted to his needs. For complex situations, that typically require many students to get focused instructions, in-depth training can be run.

Eteacher is host-independent and any program that uses mouse and keyboard can be taught.

The units can easily be designed using standard spreadsheet operations. Nearly all work of the design process for new units is done without programming.

Eteacher can handle unknown screen situations. The parameters of all entities that are needed for the control of the student are explored by a 'crawler'.

Fuzzy Control of a 3 Degree of Freedom Parallel Robot

Sergiu-Dan Stan[1], Grigore Gogu[2], Milos Manic[3], Radu Balan[1] and Ciprian Rad[1]

[1]Department of Mechatronics
Technical University of Cluj-Napoca, Cluj-Napoca, C. Daicoviciu no. 15, Romania
sergiustan@ieee.org, {radubalan+cipri_rad}@yahoo.com
[2]IFMA, Clermont-Ferrand, France
grigore.gogu@ifma.fr
[3]University of Idaho, Idaho Falls, USA
misko@ieee.org

Abstract **– Medical robotics is a relatively new branch of robotics. The paper presents an innovative user interface for the control of a 3 DOF Isoglide 3 parallel robot. The Isoglide3 robot described in (Stan, 2008), offers the superior characteristics with regards to the other parallel manipulators, such as the light weight construction. The interface presented in this paper enables the user to apply PID and fuzzy controller together. The robot interface using virtual reality was verified and tested, and results presented in MATLAB, Simulink, and SimMechanics. The fuzzy controller was numerically simulated on problems of trajectory tracking. The simulation results demonstrated the superiority of fuzzy control over traditional PID control.**

Index Terms – parallel robot, fuzzy control, ISOGLIDE3, virtual reality, nonlinear control, kinematics.

I. Introduction. Applications of Medical Robots

The artificial intelligence techniques have been recognized approach to adaptation and conversion of human knowledge into a form understandable by computers. Advanced control based on artificial intelligence techniques are typically referred to as intelligent control.

Medical parallel robots, on the other hand, have gained much popularity because of their:

- Intimate interaction with humans;
- Unstructured, irregular and dynamic environment;
- Unpredictable actions/response of operator/patient;
- Complex tasks on irregular subjects;
- Serious consequences of error or failure;
- High likelihood of emergency situations.

Nowadays, medical robots are used in orthopaedic interventions on the femur. The head, or condyle, of the femur forms the ball of the hip ball-and-socket joint.

The condyle is joined to the femur by a narrow neck where most common type of fracture occurs. Surgical repair involves stabilizing the fracture with a stainless steel fixture secured in place by screws.

The 3 DOF Isoglide3 parallel robot and its structure is shown in Fig. 1, where a mobile platform is coupled with the fixed base by three legs of PRRR type (Prismatic Revolute Revolute Revolute).

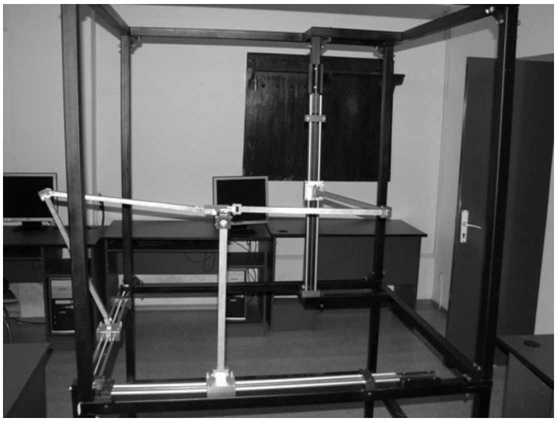


Fig. 1. 3 DOF Isoglide3 parallel robot.

The forward and inverse kinematic analysis is trivial. The three translational motions of the moving platform coincides with the motions of the three linear actuators ($x=d_1$, $y=d_2$ and $z=d_3$)

This robot architecture was also implemented and known in the literature under the name of Isoglide3-T3 (Gogu 2004, 2008), Orthogonal Tripteron (Gosselin et al. 2004), or CPM (Kim and Tsai 2002).

II. PID Control of the 3 DOF Parallel Robot

The control of the robot is implemented using a joint-based control scheme. In such a scheme, the end effector is positioned by finding the difference between the desired quantities and the actual ones expressed in the joint space [1].

The command of the robot is expressed in Cartesian coordinates of the end-effector (or mobile platform).

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_79, © Springer Science+Business Media B.V. 2010

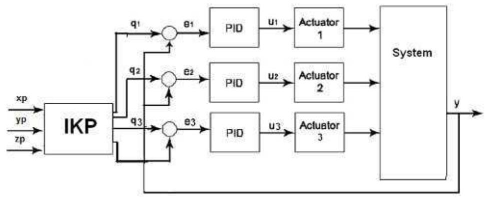


Fig. 2 Control block scheme for Isoglide3 parallel robot

Using the inverse kinematic problem, these coordinates become displacements. These displacements will become the reference for the control algorithm.

For the control of the robot, one motor element was used for each actuator of a PID control algorithm (Fig. 2). Fig. 3 presents the model of the robot using PID controllers.

The interface was designed to give a novice user an intuitive tool to control any kind of mechanical structure (serial, parallel or hybrid), requiring no programming skills at the same time.

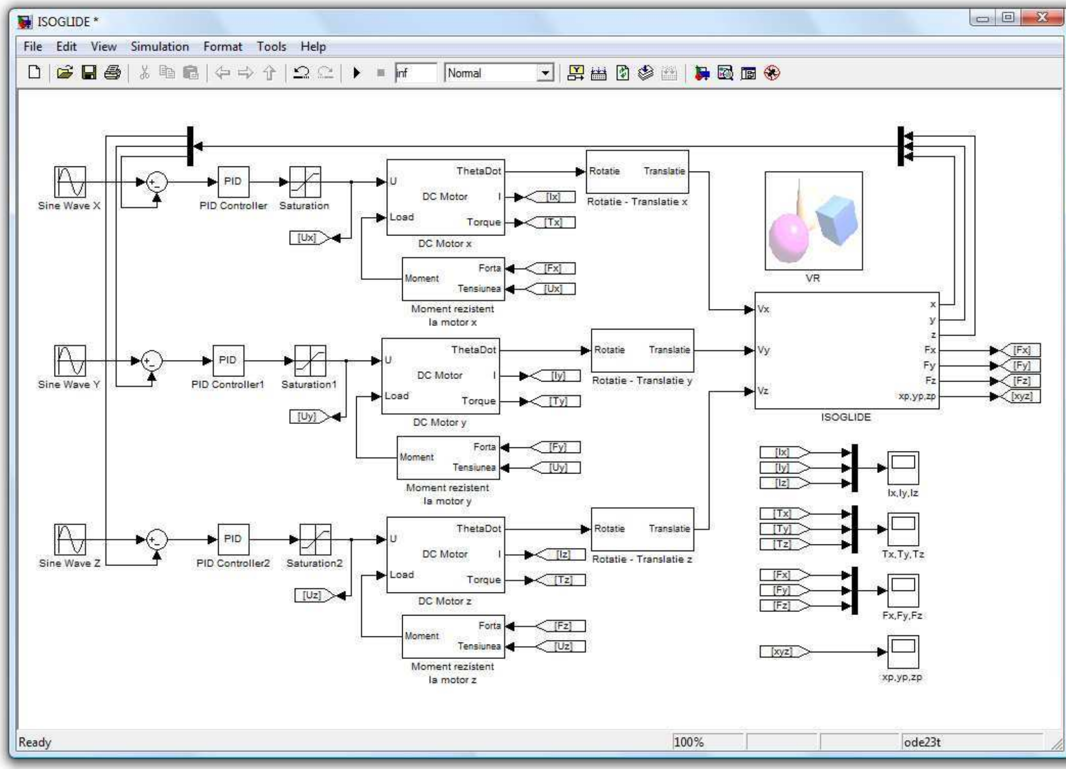


Fig. 3. Simulink model of the ISOGLIDE3 parallel robot using PID controller.

The interface is an extension of a computer based simulation that allows mimicking of a real life or potential situations with no danger imposed to humans or infrastructure.

SimMechanics models can be interfaced seamlessly with ordinary Simulink block diagrams. For example, it enables user to design the mechanical and the control system in one common environment.

In addition, Virtual Reality Toolbox for MATLAB makes more realistic rendering of bodies possible. Arbitrary virtual worlds can be designed with Virtual Reality Modeling Language (VRML), and be interfaced with the SimMechanics model. Simulation results for PID control are presented in section IV.

III. Fuzzy Control of the 3DOF Parallel Robot

Fuzzy Logic is a technique to embody human-like thinking into a control system. A fuzzy controller can be designed to emulate human deductive capabilities into in a control system. Fuzzy control incorporates ambiguous human logic into computer programs.

It suits control problems that cannot be easly represented by the mathematical models due to inherent imprecisions such are: parameter variation, unavailable or incomplete data, very complex processes. Examples can vary from optical fuzzy controllers, via control of RAID systems, to control of decision making.

Industrial electronic applications are typically based on the PID control, implemented by embedded system or PLC programming.

PID control works well for linear process, but does not perform satisfactory when it comes to non-linear proceses. The idea of fuzzy control of the parallel robots was tested and demonstrated in ANFIS Matlab environment tool [16].

Design of a fuzzy controller requires more design decisions than usual, for example regarding rule base, inference engine, defuzzification, and data pre- and post processing.

Fig. 4 presents the model of the robot using fuzzy controllers.

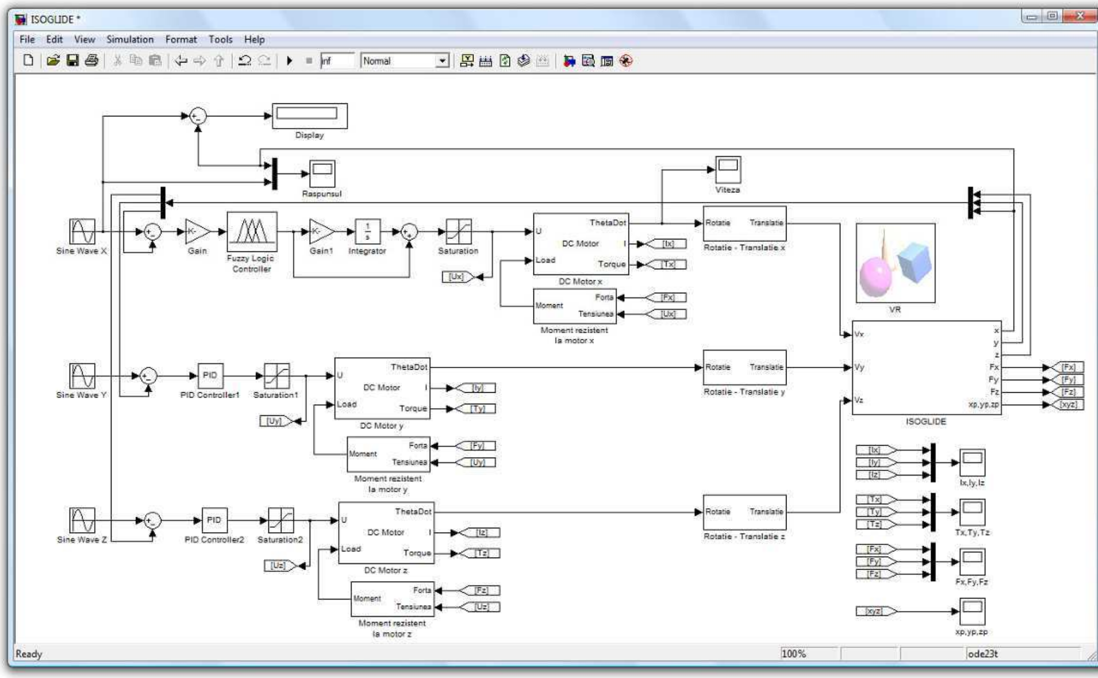


Fig. 4. Simulink model of the ISOGLIDE3 parallel robot using fuzzy controller.

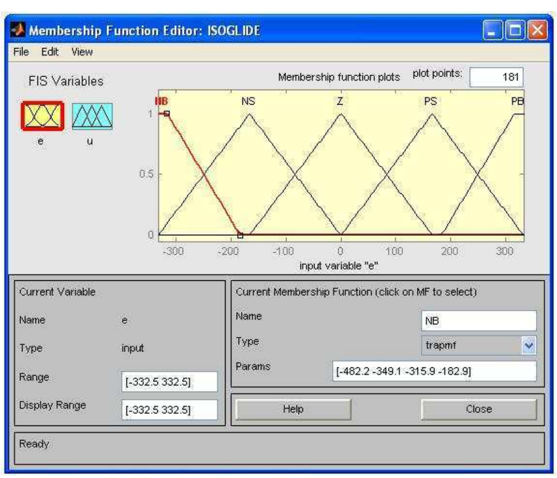


Fig. 5. Membership function block for input variable

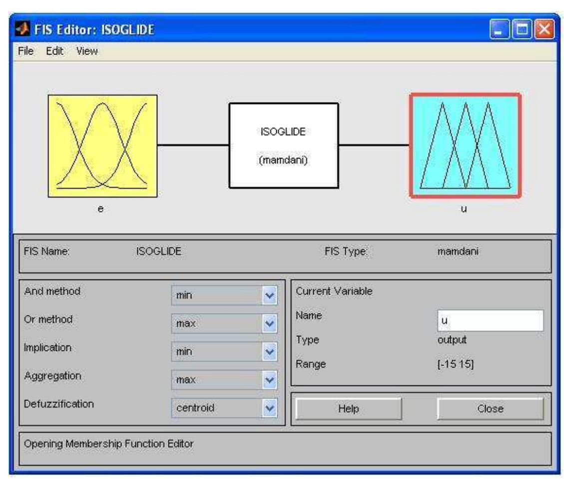


Fig. 6. Fuzzy block scheme for 3DOF ISOGLIDE parallel robot

The first block inside the controller is the fuzzification module, which converts each piece of input data to degrees of membership by a lookup in one or several membership functions (Fig. 5 and Fig. 6).

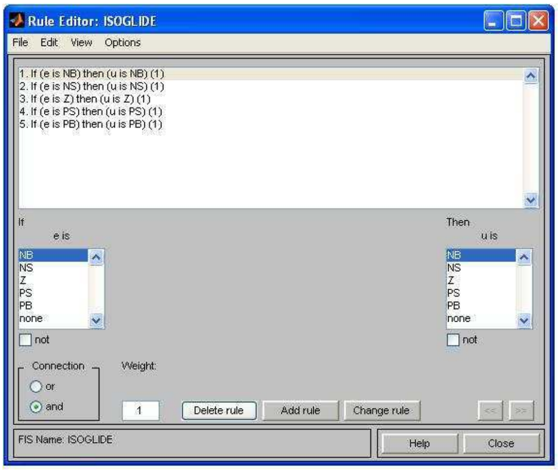


Fig. 7. Rule based editor from Matlab/Simulink

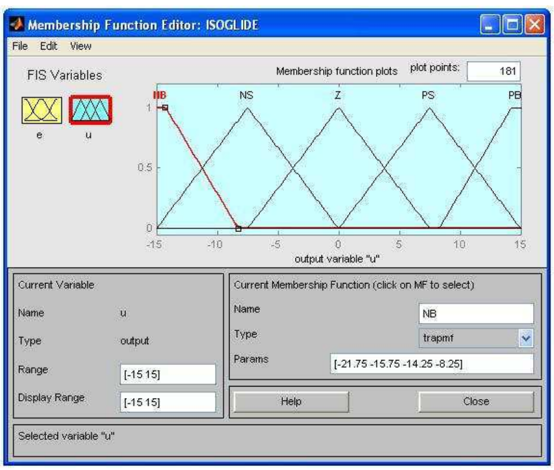


Fig. 8. Membership function block for output variable

Design of efficient control for dynamic system usually requires a prior knowledge of the process to be controlled. The first step is to obtain mathematical model for the process, actuator and the controller.

Sometimes, the accurate modeling is very difficult or nearly impossible in which cases Fuzzy Logic Controllers (FLC) have been successfully applied.

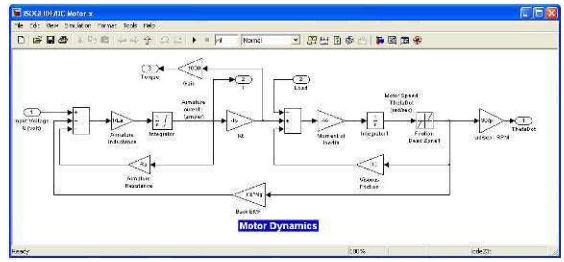


Fig. 9. Simulink block scheme of the motor dynamics

Fig. 9 presents the Simulink block scheme for the motor dynamics, while Table 1 presents the DC motor parameters used in the simulation.

TABLE 1. DC MOTOR PARAMETERS

Armature resistance, Ra [ohm]	0.334
Armature Inductance, La [H]	0.000085
Torque Coefficient, Kt [N*m/A]	0.0194
Back EMF Coefficient, Kb [V*s/rad]	0.0194
Gear Ratio, Ng [rad/rad]	4.8
Moment of Inertia of Armature, Ja [kg*m^2]	0.00000676
Moment of Inertia of Gear/Load, Jg [kg*m^2]	0.0000156175
Coefficient of viscous friction at armature, Ba [N*m*s/rad]	0.000062
Coefficient of viscous friction at gear, Bg [N*m*s/rad]	0.00001
Current saturation coefficient, Isat [A]	4
Friction dead zone coefficient, dz [N*m*s/rad]	0.1

Fig. 10 presents the Simulink model for resistant moment of actuators.

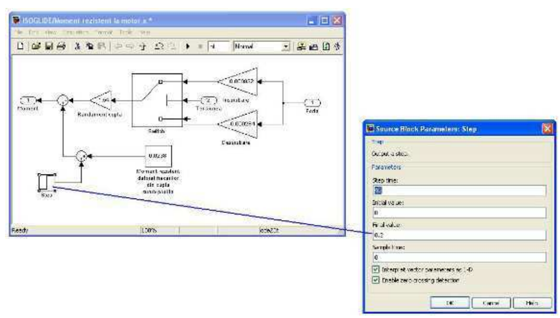


Fig. 10. Simulink model for resistant moment of actuators

There is no design procedure in fuzzy control such as root-locus design, frequency response design, pole placement design, because the rules are often nonlinear.

IV. Numerical Simulations of the 3Degrees of Freedom Parallel Robot

The path is defined as sequence of robot configurations in a particular order without regard for timing of these configurations while trajectory is concerned about when each part of the path must be obtained thus specifying timing.

The first tests on the prototype encourage the direction of the research: the chosen control algorithms emphasize the peculiar characteristics of the parallel architecture and, in particular, the good dynamic performance due to the limited moving masses, and advantageous robot behaviour.

The interface is based on a virtual reality approach in order to provide the user with an interactive 3D graphical representation of the parallel robot.

The sample trajectory of the end-effector is chosen to be a circular path with the radius of 0.4 meters and its center is O(0 ,0,0). The end-effector path is shown in Fig. 12.

The desired force obtained from the actuators to move the end-effector of the Isoglide3 parallel robot along the desired trajectory is shown in figure 13.

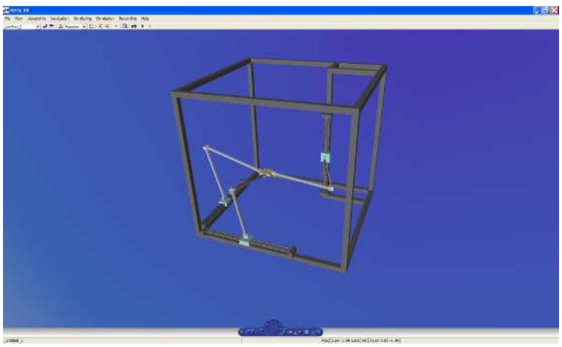


Fig. 11. ISOGLIDE3 virtual reality robot interface.

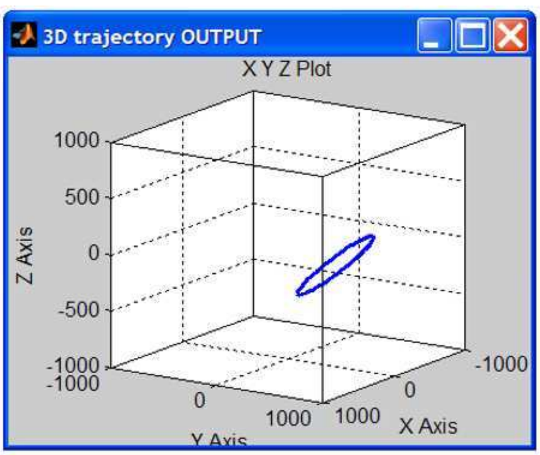


Fig. 12. End-effector path for the 3D circular trajectory.

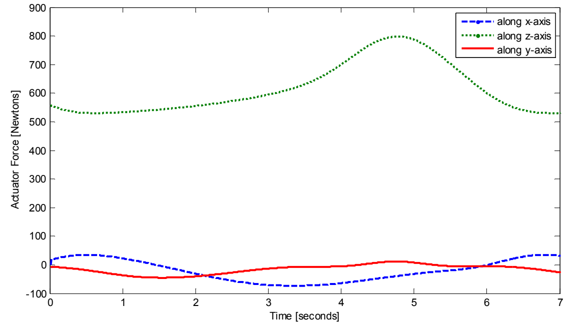


Fig. 13. The desired force obtained from the actuators.

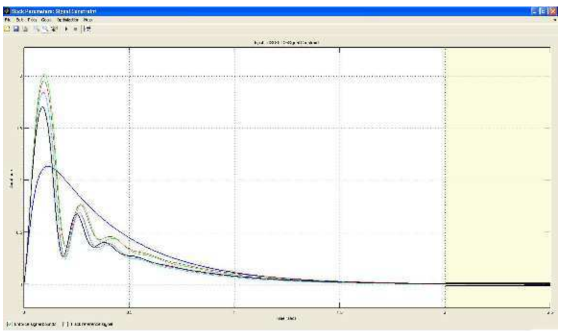


Fig. 14. PID optimization

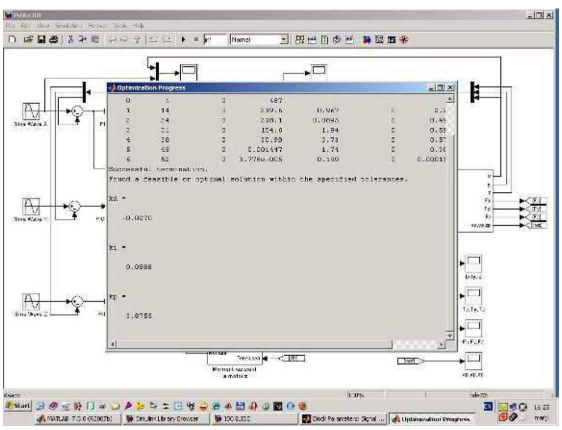


Fig. 15. PID optimization results

The controller parameters *kp, ki* and *kd* were optimized for a given trajectory and a maximum error, using the block Signal Constraint from the Simulink Response Optimization toolbox (see Fig. 8). After optimization it was found the following PID parameters: *kp=3.8755, ki=9.0888* and *kd=-0.0270.*

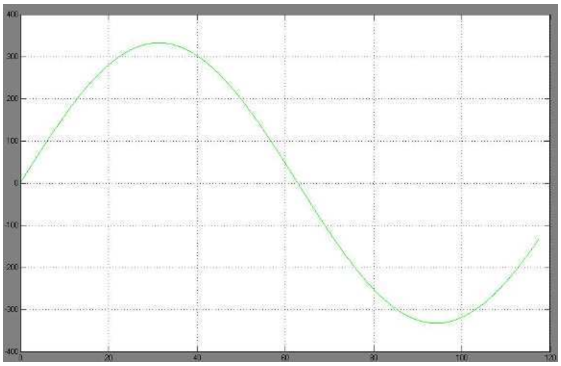


Fig. 16. Fuzzy control results for sinus tracking error

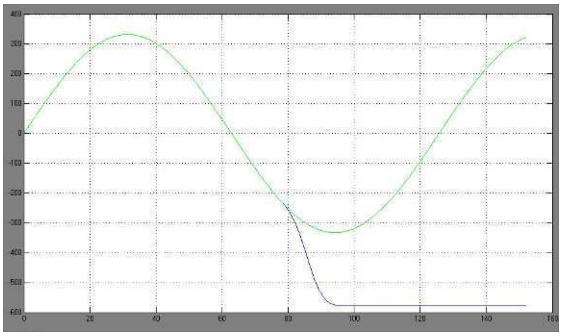


Fig. 17. PID control results for tracking error

Fig. 16 & 17 presents the simulation results, as one can see the fuzzy control presents better performances as PID. In a fuzzy controller the data passes through a preprocessing block, a controller, and a postprocessing block.

The steps that were applied in the design procedure based on PID and fuzzy control were:

1. Tune a PID controller
2. Replaced it with a linear fuzzy controller
3. Fine-tune it.

Preprocessing consists of a linear scaling as well as quantization in case the membership functions are discretised (vectors); if not, the membership of the input can just be looked up in an appropriate function.

IV. Conclusions

The applications of micro machine, mechatronics and robotics in medicine are widely spread in clinical use. There are a number of areas in which robotic technology has the potential to contribute positively to the provision of healthcare. The paper presents a fuzzy and PID control for the 3 DOF Isoglide3 parallel robot. Fuzzy controllers are being used in various control schemes. The controller is here a fuzzy controller, and it replaces a conventional controller, a PID (proportional-integral-derivative) controller and proves to have better results.

Acknowledgment

Support for this work was provided through the CNMP under research grant PARTENERIATE 3280, "Complex mechatronics systems for medical applications".

References

[1] C. Gosselin. "Determination of the workspace of 6-d.o.f. parallel manipulators". *ASME Journal of Mechanical Design*, 112:331–336, 1990.

[2] J. P. Merlet. "Determination of the orientation workspace of parallel manipulators". *Journal of intelligent and robotic systems*, 13:143–160, 1995.

[3] S. Stan, V. Maties, R. Balan, „Optimal Design of a 2 DOF Micro Parallel Robot Using Genetic Algorithms", Proceedings of the 2007 IEEE-ICIT 2007, IEEE *International Conference on Integration Technology,* March 20 - 24, 2007, Shenzhen, China, p. 719-724. IEEE Catalog Number: 07EX1735, ISBN: 1-4244-1091-6, ISBN: 1-4244-1092-4.

[4] Gogu, G., Structural synthesis of fully-isotropic translational parallel robots via theory of linear transformations, European Journal of Mechanics / A –Solids, vol. 23, pp. 1021-1039, 2004.

[5] Gogu, G., Structural synthesis of parallel robots, Part 1: Methodology, Springer, 2008.

[6] Stan, S.-D, Manic, M., Mătieş, M., Bălan, R., "Evolutionary Approach to Optimal Design of 3 DOF Translation Exoskeleton and Medical Parallel Robots", HSI 2008, IEEE Conference on Human System Interaction, Krakow, Poland, May 25-27, 2008.

[7] Stan, S.-D, Manic, M., Mătieş, M., Bălan, R., "Kinematics Analysis, Design, and Control of an Isoglide3 Parallel Robot (IG3PR)", IECON 2008, The 34th Annual Conference of the IEEE Industrial Electronics Society, Orlando, USA, November 10-13, 2008.

[8] Kim, S.H., Tsai, L-W., "Evaluation of a cartesian parallel manipulator", In: Lenarčič J, Thomas F (eds) Advances in robot kinematics. Kluwer Academic Publishers, pp 21-28, 2002.

[9] Craig, J., Robotics, Mechanics and Control, Addison-Wesley Publishing Company, USA, 1995.

[10] Isermann, R., Mechatronic Systems, Fundamentals, Springer, Londra, Anglia, 2005.

[11] Zadeh, L.A. (1968). "Fuzzy algorithms". Information and Control 12 (2): 94–102. doi:10.1016/S0019-9958(68)90211-8. ISSN 0019-9958.

[12] Stan, S.D., Manic, M., Maties, V., and Balan, R., A Novel Virtual Reality Robot Interface for Isoglide3 Parallel Robot, Lecture Notes in Computer Science, (LNCS)/Springer, C. Xiong et al. (Eds.): ICIRA 2008, Part I, LNAI 5314, pp. 1265–1275, © Springer-Verlag Berlin Heidelberg 2008.

[13] Zadeh, L.A. (1965). "Fuzzy sets". Information and Control 8 (3): 338--353. doi:10.1016/S0019-9958(65)90241-X. ISSN 0019-9958.

[14] Cox, Earl (1994). The fuzzy systems handbook: a practitioner's guide to building, using, maintaining fuzzy systems. Boston: AP Professional. ISBN 0-12-194270-8.

[15] Klir, George J.; Folger, Tina A. (1988). Fuzzy sets, uncertainty, and information. Englewood Cliffs, N.J: Prentice Hall. ISBN 0-13-345984-5.

[16] ANFIS Matlab, URL from Oct.08, http://www.mathworks.com/access/helpdesk/help/toolbox/fuzzy/.

[17] Golden, L., Manic, M., "Optical Fuzzy Logic for Low-Power Satellite Controller", IECON08, The 34th Annual Conference of the IEEE Industrial Electronics Society, Orlando, Florida, Nov. 10-13, 2008.

[18] McCarty, K ., Manic, M., "Contextual Fuzzy Type-2 Hierarchies for Decision Trees (CoFuH-DT) – An Accelerated Data Mining Technique" , HSI 2008, IEEE Conference on Human System Interaction, Krakow, Poland, May 25-27, 2008.

[19] Guillermo N., Manic, M., "Fuzzy Control of Sparing in Disk Arrays", 12th IEEE Conference on Emerging Technologies and Factory Automation, ETFA'07, Sep. 25-28, 2007, Patras, Greece.

eduPAL: An Online Service to Create, Manage and Evaluate Educational Projects

Jhilmil Jain, Malena Mesarina and Mohamed Dekhil
Hewlett-Packard Laboratories
1501 Page Mill Road
Palo Alto, CA 94304 USA

Abstract- **In this paper we describe an online service that allows students to create research-based projects assignments very easily from different sources (desktop, laptop, mobile devices) and in different formats (text, still images, audio, video, annotated media). It also provides features to allow students to not only manage projects collaboratively, but also provides customizable templates to automatically generate well-paginated reports with several layout options. In addition, it allows teachers to automatically create project assignments as templates, and uses the class roster to allow the teacher to assign the projects to individual or groups of students, and monitor the progress and assignment status for any student or group of students in real time. The service uses a combination of natural language processing algorithms and data analytics to provide semi-automatic tracking, evaluation and dynamic visualization of the students' progress in projects assignments over time.**

I. Introduction and Motivation

Project-based instruction (PBI) is an example of an active teaching methodology that engages students by allowing them to perform individual or group investigation of a topic to develop valuable research skills [1]. While, PBI comes in many flavors (such as problem solving, model building etc.), in this paper, we are concentrating on project assignments that focus on conducting investigation or research activities.

Research shows [2, 3] that PBI should be an integral and critical part of education as it motivates students, helps them to better understand the subject matter, teaches them to acquire and apply concepts, and finally it has the potential to improve competence in thinking because students need to formulate plans, track progress, and evaluate solutions for a given topic.

Though PBI has many advantages, the adoption is not as widespread due to the lack of software and services for both students and teachers. We have identified primarily four problems in this space:

1) Tools for project creation, organization and management are not integrated. Currently there are barely any tools that help students easily collect information from a variety of different sources (Web, devices) and in different formats (text, audio, video, images etc.). Students have to create, manage, organize and share these projects manually or by using tools that are not well integrated.

2) Teachers assignment of projects to student groups is done manually and is time-consuming. Teachers typically need to assign varying levels of the same project to different classes or groups. For example, the teacher may create two versions of the same project (e.g. "Solar System") for two different classes/groups – regular and advanced program (AP).

3) Project tracking and evaluation is typically done manually by teachers. Teachers have no tools available to automatically track the progress and evaluate the status of student projects.

4) Visualization of student progress in projects is static and non-configurable. The tools that are available for visualizing the analytics of the projects (i.e. time taken, websites visited etc.) are static and not configurable (e.g. Excel sheet that shows only time taken to complete a task).

II. User Study To Understand The Problem Space

We conducted a small user study to understand the practical challenges faced both by students and teachers while completing research-based projects. We also wanted to learn about the tools, strategies and resources used by both groups. We interviewed 6 teachers and 15 students (from grades 6, 7 and 8) from three schools located in Oakland, East Palo Alto, and Fresno in California, USA. The teachers were primarily responsible for teaching the following subjects: social sciences, language arts, mathematics, and science. The following is a summary of our findings from the study:

1) *Importance of software tools to conduct research:* We found that with the proliferation of the Internet and the availability of credible resources on the Web, online information gathering has overtaken other traditional mediums of research, such as library books, printed encyclopedias and magazines. Students that we observed usually printed web-pages of interest, and then highlighted relevant sections. They struggled with keeping a log of interesting websites that were visited and felt that they were "*killing the trees*" by printing unnecessary pages.

2) *Importance of the use of mobile phones:* Most students that we interviewed owned cell phones and used it primarily for texting and taking pictures.

"With my cell phone I can text something like 'make sure you read in 15 minutes'. That would help a lot because I forget. I also text my friends asking them what the assignment was so that way I wouldn't miss it".

When asked if they would prefer a laptop over a cell phone to conduct project based research most said they would prefer a cell phone since:

"[Using the cell phone] you can call, text, Internet, blog, photos, audio, video, music. That's all in one and it's so tiny and fits in everything"

When we interviewed teachers, they mentioned that cell phones are usually not allowed in the classrooms as students get distracted and don't use it for learning purposes. But if the

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_80, © Springer Science+Business Media B.V. 2010

teachers could use technology to integrate mobile phones for learning, they feel these would engage the students in active learning exercises.

3) *Importance of automated evaluation support:* Teachers typically provide checkpoints and deadlines of project sections to be completed by a certain date. Since most teachers usually have a high workload, and it is becoming increasingly difficult for them to provide individual attention to each student, thus there are a number of opportunities to use technology for automating progress tracking and evaluation of projects.

III. SOLUTION

A. Project template creation used by the teachers

The selection of a topic is the first step that a teacher takes when assigning a research project to students. She may also want to partition the topic into sub-topics to emphasize particular areas. *eduPAL* helps the teacher in this project creation stage by automatically suggesting related concepts that the teacher can choose as sub-topics. In addition, this set of related concepts is also used by *eduPAL* to calculate metrics that will aid the teacher in evaluating a student's progress in completing the project (refer to part C of this section for details related to metrics).

In order to generate this set of concepts, we developed an algorithm that extracts domain specific concepts from web-pages that are related to the main topic. The algorithm has two stages-search and extraction. In the search stage, the algorithm takes the name of the topic as a seed and runs a Google search. It uses the top five web-pages returned by Google as an input to the extraction stage. In the extraction stage, the algorithm applies a tri-gram part-of-speech tagger [4] that uses a general domain corpus (Penn Treebank) as the training set, to classify the words in the web-pages (i.e., verb, noun, etc). The unclassified concepts have a high likelihood of belonging to the set of concepts that are specific to the topic of interest. The intuition behind the algorithm is that a general purpose tagger like the Penn Treebank, which uses the Wall Street Journal as its trained corpus, will not be able to tag words that are specific to a narrower domain, like the seeded topic.

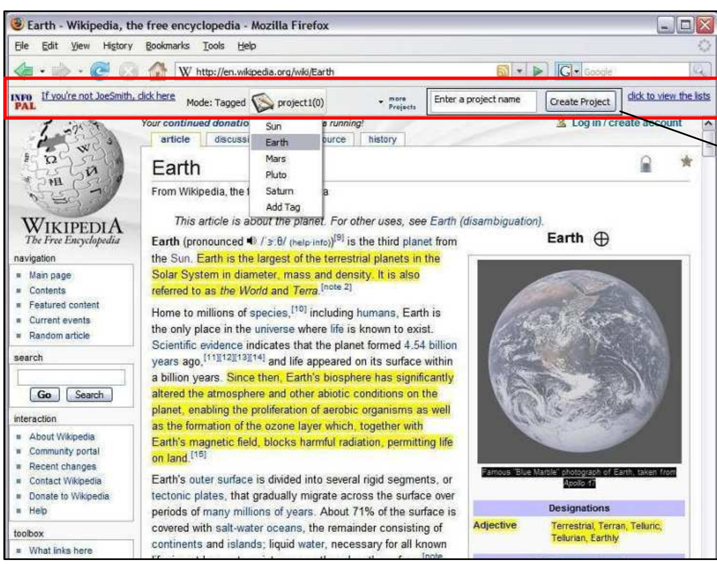
After generating concepts for all the five web-pages, our algorithm ranks and selects the first 50 most frequently used words. The algorithm also assigns a weight to each concept, which is equal to its frequency count divided by the sum of the frequency count of all 50 concepts. Once the concepts are generated, the teacher can select sub-topics from this set of concepts. The sub-topics are then used in a second round of concepts generation to find the related concepts associated with each sub-topic.

For purposes of illustration, let us assume that the teacher has created a project template called the "Solar System" and has assigned the following sub-topics for it: "Sun", "Earth", "Mars", "Pluto", and "Saturn". This interaction is carried out by the teacher in the backend. As soon as the template is released, it is available in the drop down box of the *eduPAL* toolbar (Fig 1.b) for use by the student.

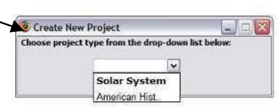
B. eduPAL user interface used by students to create, collect, manage, organize, share, and publish educational projects

In this section we will give an overview of the user interface that is used by the students to create and manage projects. Fig. 1a shows the Firefox browser with the *eduPAL* toolbar (refer to the red box) installed on it. To create a project, the student types in a suitable name in the text field (e.g. project1), and clicks on the "Create Project" button. Fig. 1b shows the available project templates that have been assigned by the respective teachers to the student. In our case, let's assume *the* student chooses the type "Solar System" for the project. As soon as the project is created, a customized button *project1(0)* appears on the toolbar. The first part of the button states the name of the project and the count on the parenthesis indicate the number of informational items added to the project.

The toolbar allows the student to simply highlight the relevant text and images on a webpage, and click on the project name (e.g. project1) to add the information to it. As the student clicks on the project name, a list of project sub-topics (as defined by the teacher) appear as a drop down menu.



a) Collect information from the web-page



b) This drop down box is populated as soon as the teacher releases a project template

Figure 1 (a and b): Demonstration of student interaction with the eduPAL toolbar using a browser

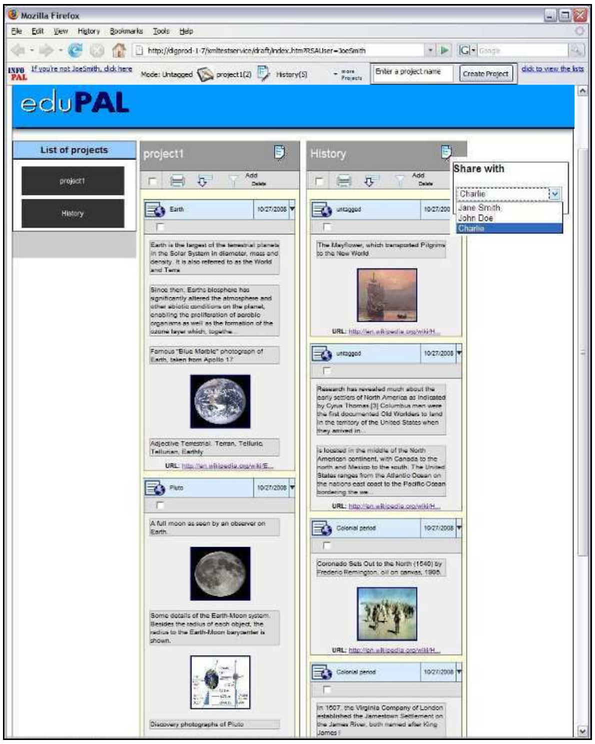


Figure 2: Manage, organize and share projects

In Fig 1a, it is shown how the student can add information to the sub-topic “Earth” of the project named “project1” (this project is of the type “Solar System”). The toolbar enables the student to add information to the predefined sub-topics chosen by the teacher. Fig. 2 shows the *eduPAL* user interface that allows the students to view, manage, organize, share and print the projects (this webpage can be access by clicking the link “click to view the lists of projects” on the toolbar - see Fig .1).

Items (i.e. collected information) can be moved between sub-topics in a particular project, or between two projects by simply dragging and dropping them. Projects can be shared between students, by right clicking on the project name and choosing the names of the relevant people. The names of the students are pre-populated from the class roster. Students can add their own freehand notes to the projects as well. Fig. 3 shows the mobile client interface that students can use to add audio, video, images etc. from their phone to the projects. We are currently implementing capabilities such that students can use SMS[1] and MMS[2] to add text and multimedia respectively to the projects.

[1] Short Messaging Service
[2] Multimedia Messaging Service

Project reports can be automatically created using predefined templates provided by the teacher, or these can be customized by the students. Students can choose to print the whole project, or can “check mark” parts of the project to be printed. Fig. 4 shows an example of a “well-paginated” report that has been automatically generated by *eduPAL* for the project named “project1”.

C. Metrics used to measure and track project progress

eduPAL provides three metrics to aid the teacher in evaluating the progress of the students in the completion of a project. These metrics provide the teacher with numerical measures on how the students are progressing in the data collection part of the project.

We define three metrics: relevance, coverage and uniqueness. Relevance measures how relevant the articles collected by a student are with respect to the topic/sub-topic. Coverage is a measure of how exhaustive is the research done by students. It measures the percentage of topic/sub-topic related concepts that have been found in the collected articles. Uniqueness is a measure of how much effort the students have put in collecting distinct facts about a topic/sub-topic. All things being equal, collecting articles that capture different facts is better than collecting articles with the same facts. Next we will briefly explain how these metrics are computed.

1) *Relevance*: *eduPAL* assigns to each collected article a relevance value that depends on the number and the weight of the related concepts that appear in the article. The related concepts are the set of concepts found by eduPAL during the project creation process.

The relevance of an article is the average weight of all the related concepts found in the article and has a range between 0 and 1. To compute the project relevance, we average the relevance values of all articles.

2) *Coverage*: *eduPAL* assigns to a topic or sub-topic, a coverage value that measures what percentage of the generated concepts for that topic or sub-topic has been covered by all the articles collected for it. To compute the project’s coverage, we take the average of the coverage values of all sub-topics.

3) *Uniqueness*: *eduPAL* assigns a uniqueness value U_{ij} to each pair of articles for a topic/sub-topic, which measures how different the two articles are from each other with respect to the related concepts generated for the topic/sub-topic. In other words, do the articles use the same concepts or different ones?

We use the cosine similarity equation measure shown in Equation (1) as a measure of the uniqueness U_{ij} between two articles. The binary vectors, $\overline{D_i}$ and $\overline{D_j}$, indicate whether a generated concept is present (‘1’) or not in the articles (‘0’), and the cosine similarity results in a value between -1 and 1, where -1 indicates that they don’t share any concepts and 1 that they share all the concepts.

Figure 3: Mobile client to capture multimedia from handhelds

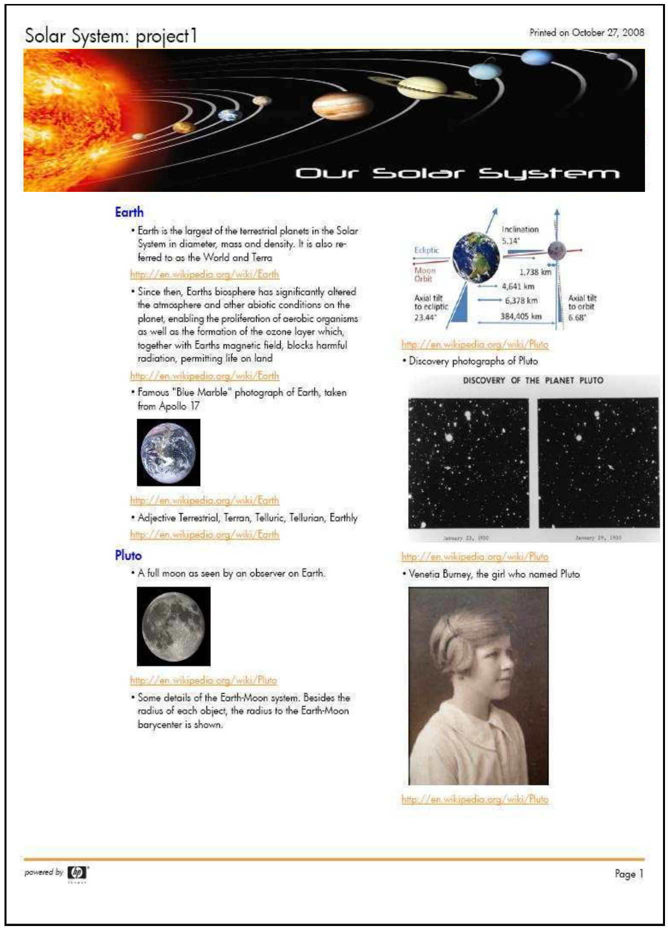


Figure 4: Print quality automatic report generation

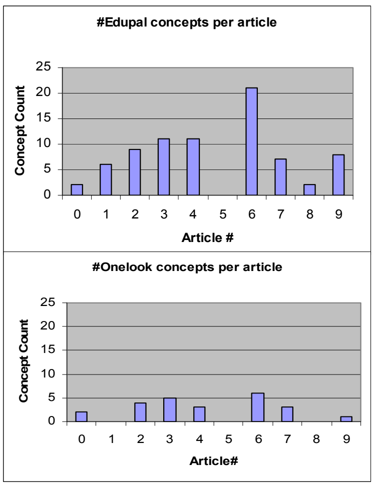


Figure 5: Comparison of distinct concepts extract by eduPAL and OneLook

eduPAL computes the uniqueness of topic/subtopic by taking the average of all the U_{ij} pairs for each subtopic. To compute the uniqueness of the project, we average the uniqueness values of all sub-topics

$$\cos(D_i, D_j) = \frac{D_i \bullet D_j}{|D_i||D_j|} = U_{ij} \tag{1}$$

Experimentation and Results

While the metrics implementation in *eduPAL* is currently under development, we provide some preliminary experimental results of our concept generation algorithm that validate the intuition behind the approach. We compared our results with another concept generation tool, the OneLook Reverse dictionary [6], which is publicly available on the web.

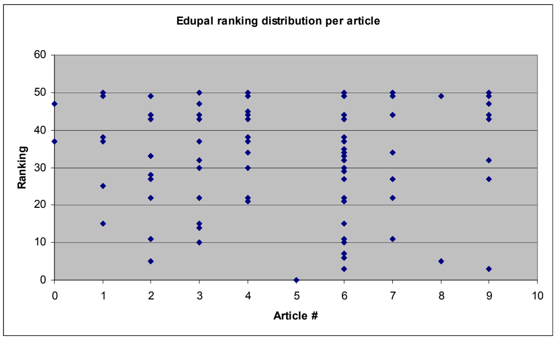


Figure 6: *eduPAL* ranking distribution

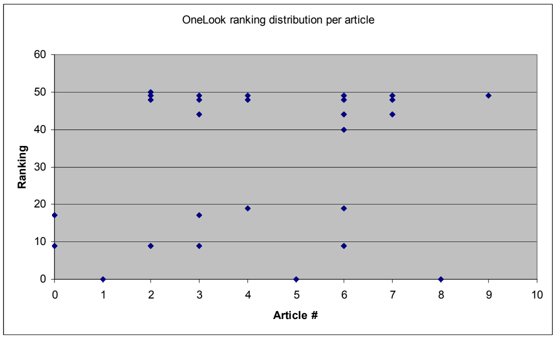


Figure 7: OneLook ranking distribution

As a seed topic to our algorithm, we use the topic name "Solar System", and obtained 50 ranked related concepts. We entered the same topic name to the OneLook tool, and also obtained their first 50 related concepts. The concepts returned by OneLook are ranked by relatedness (as stated in on OneLook website).

We compared our concepts with the ones returned by OneLook using two metrics: 1) how many different concepts were used per article, and 2) the distribution of ranking of the words per article. The first metric can be used to compare which approach produces concepts that are most likely to be found in an article, while the second metric can be used to compare which approach produces the highest ranked (more relevant concepts) to be found in an article.

The articles chosen for the comparisons where the next 10 web-pages that were initially found by conducting a Google search for the keywords "Solar System". The first 5 web-pages returned by the Google search were excluded since these were used as the training set for our algorithm.

Figure 5 shows the comparison of results for the number of distinct concepts generated per article for *eduPAL* and OneLook. We can see that more *eduPAL* generated concepts are found in articles than OneLook concepts. Therefore, we conclude that using *eduPAL* generated concepts to compute our metrics, would give a more accurate indication of whether an article chosen by student contains relevant information.

Figure 6 and Figure 7 show the concept ranking distribution for *eduPAL* and OneLook respectively. We can see that on average all articles contain concepts from *eduPAL* whose ranking is evenly distributed (Fig 7). In contrast, all the articles show that the OneLook concepts found in the articles are at either extreme of the OneLook ranking. We interpret from this result that because the ranking distribution of *eduPAL* concepts found in the articles is evenly distributed vis-à-vis the bipolar distribution of OneLook, using *eduPAL* concepts to compute our metrics will give a more accurate measure of the relatedness of the article. For example, for articles 2, 3, 4, 6 and 8, if OneLook concepts were to be used for our metrics, all the words with ranking between 20 and 40 that *eduPAL* includes, would have been ignored.

D. Visualization of results

Student progress in the projects can be semi-automatically tracked using the metrics detailed above and can be visualized dynamically in real time (see figures 8a-8c). Visualizations can be created per student, per group or per class. Progress along several dimensions such as project sub-topics, websites used, time spent on the assignment, and others can be visualized.

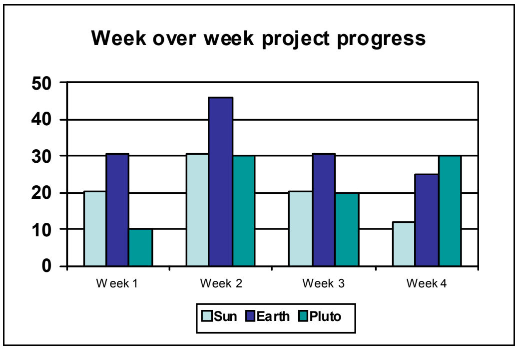


Figure 8a: Example of student progress week over week

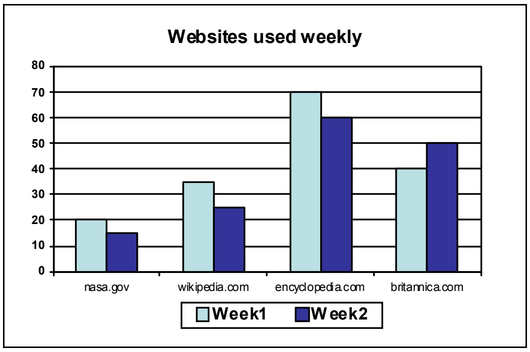


Figure 8b: Websites visited weekly for data collection

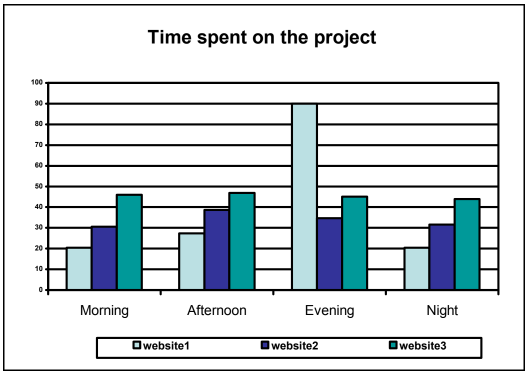


Figure 8c: Time spent working on the project

IV. RELATED WORK

Blackboard [7] and eCollege [8] are the two main competitors in the space of online web-based educational software. Blackboard develops software that enables schools to create web-based learning programs and communities. The company's *Academic Suite* and *Commerce Suite* connect teachers, students, parents, and administrators via the Web, enabling web-based assignments, class web sites, and online collaboration with classmates. The software also assists instructors with course administration and includes a content management system for creating and managing digital course content. Blackboard however lacks the "data collection" gathering capabilities and the automatic "student progress analytics" that our solution proposes. Using Blackboard, a student can only upload documents but can't capture data through the different channels (i.e. mobile phone, mouse selection of text and other media) and various modalities (text, video, pictures, voice, etc) that Edupal provides. In addition, Blackboard only provides statistical results (i.e. mean, median, standard deviation) and charts given grades enter by the teacher. There is no use of natural language processing techniques or data mining techniques to provide an automatic assessment of progress to the teacher, as we do in *eduPAL*.

eCollege also provides an educational software system which they call the *Course Management System* that allows course authoring, delivery and interactions with students. Like Blackboard it lacks the "data gathering" and automatic "student progress analytics" that *eduPAL* provides.

There has been considerable research done to automatically evaluate computer programs by systems like CodeLab [9, 10], ASSYST [11], TRAKLA [12], RoboProf [13], BOSS [14], CourseMaker [15], but to the best of our understanding there is no system available that aids in the evaluation of research based project assignments containing multimodal data such as text, images, video etc.

V. CONCLUSIONS

eduPAL provides a number of advantages over existing software and services. For students, it allows them to:

1. Collect, manage and organize information for a project from different sources (Web, mobile) and in different formats (structured text, unstructured text (scribbles), still images, audio, video, annotated media)
2. Create "print quality" project reports using templates provided by the teachers, or templates that can be customized by the students, thereby reducing the effort spent on output creation.

For teachers (and parents), it allows them to:

1. Create project assignments electronically
2. Track the progress per student, or per group or per class. The progress parameters and timeline are customizable. Timeline can consist of few days to weeks to month. Parameters can consist of websites visited, time spent, contribution per person (if it's a team project), percentage of assignment completed etc.
3. Evaluate student projects semi-automatically by providing the teacher with a range of evaluation metrics

ACKNOWLEDGMENT

We would like to thanks Ismail Ari for organizing the interviews at the school in Oakland, CA. Brant Jones and Erik Geelhoed for organizing and conducting the interviews at the school in Fresno, CA.

REFERENCES

[1] B. Barron, "Doing with understanding: Lessons from research on problem- and project-based learning," *Journal of the Learning Sciences*. 7(3&4), pp. 271-311.

[2] P.C. Blumenfeld, E. Soloway, R.W. Marx, J.S. Krajcik, M. Guzdial, A. Palinscar, "Motivating project-based learning: Sustaining the doing, supporting the learning," *Educational Psychologist*, 26(3 & 4), pp. 369-398.

[3] T. Bredderman, "Effects of activity-based elementary science on student outcomes: A quantitative synthesis," *Review of Educational Research*, 53, pp. 499-518.

[4] Natural Language ToolKit. Available at: http://nltk.sourceforge.net/index.php/Main_Page

[5] C. D. Manning and H. Schutze, "Foundations of Statistical Natural Language Processing," *The MIT Press*, 1999, pp. 300

[6] OneLook Reverse Dictionary. Available at: http://www.onelook.com/reverse-dictionary.shtml

[7] Blackboard. Available at: www.blackboard.com

[8] eCollege. Available at: www.ecollege.com

[9] D. Arnow and O. Barshay, "On-line programming examinations using WebToTeach," *In Proceedings of the 4th Annual SIGCSE/SIGCUE Conference on Innovation and Technology in Computer Science Education*, Krakow, Poland, June 27 – 30, pp. 21–24.

[10] D. Arnow and O. Barshay, "WebToTeach: An Interactive Focused Programming Exercise System," *In Proceedings of FIE '99,* San Juan, Puerto Rico, November 1999, IEEE Press, Session 12a9.

[11] D. Jackson and M. Usher, "Grading student programs using ASSYST," *In Proceedings of the 28th SIGCSE Technical Symposium on Computer Science Education,* San Jose, CA, Feb. 27–March 1, pp. 335–339.

[12] J. Hyvönen, and L. Malmi, "TRAKLA – A system for teaching algorithms using email and a graphical editor," *In Proceedings of the HYPERMEDIA Conference (Vaasa),* pp. 141–147.

[13] C. Daly, "RoboProf and an introductory computer programming course," *In Proceedings of the 4th Annual SIGCSE/SIGCUE Conference on Innovation and Technology in Computer Science Education,* Krakow, June 27 – 30, pp. 155 – 158.

[14] M. Luck and M.S. Joy, "A secure on-line submission system,". *Software Practices and Experiences*, 29(8), pp.721–740.

[15] C. A. Higgins, G. Gray, P. Symeonidis, A. Tsintsifas, "Automated Assessment and Experiences of Teaching Programming," *Journal on Educational Resources in Computing (JERIC).* 5(3), September 2005, pp. 1531-4278.

Integrating Web 2.0 Technologies in Undergraduate Teaching: Experiences with a Wiki Implementation

Ilias Karasavvidis
Department of Preschool Education
School of the Humanities
University of Thessaly
ikaras@uth.gr

Abstract - **The present paper presents our experiences with a wiki implementation in undergraduate education. More specifically, the paper address the issue of collaboration in the process of hypertext creation and examines the rate and dominant patterns of student collaboration. Fifty students who enrolled in an undergraduate course on learning with ICT participated in the study. One of the compulsory assignments involved the use of a wiki system for the collaborative construction of a textual artifact which aimed to facilitate the shared development of meaning. Student collaboration was operationalized in terms of joint revisions. Results indicated that the levels of student collaboration were low and that in the few instances of joint revision the predominant pattern was a cumulative one. The paper is concluded with a discussion of the importance of scaffolds for the effective integration of wikis in higher education.**

Keywords – Wiki, Collaborative Meaning Making, Joint Revision, Web 2.0, Pedagogical Activity

I. Introduction

The potential of e-learning to improve the quality of student learning in higher education is undisputed. Still, e-learning educators face major challenges which are both multifaceted and complex ([1], [2]). One of the most important challenges educators face is related to pedagogy. For instance, typical e-learning approaches tend to replicate traditional transmissionist practices which are fairly didactic and are mostly based on behaviorist conceptions of learning ([3], [4], [5]). The implementation of a different, more constructivist pedagogy requires new tasks, activities, and practices. One of the challenges for e-learning and blended learning is to promote more constructivist pedagogies.

Web 2.0 technologies present important opportunities for changing learning practices for tertiary education. Wiki systems in particular can help design and implement a different, more constructivist pedagogy compared to the typical content-delivery which characterizes much of contemporary e-learning ([6]).

A. Wiki Systems

In line with the widespread adoption of Web 2.0 technologies, wikis are gaining stunning acceptance. "WikiWiki" is a Hawaiian word that means quick. According to [7], a wiki is a "*freely expandable collection of interlinked webpages, a hypertext system for storing and modifying information – a database, where each page is easily edited by any user with a forms-capable Web browser client*" (p. 14). Essentially, a wiki system constitutes a collection of webpages with information which can in principle be edited by anyone. Users can add content to existing pages, restructure the information on existing pages as well as create new pages and create links between existing and new pages. The fact that everyone can create, edit and delete a page means that wikis have a very open-ended nature which can be characterized by a democratic participation pattern. WikiWikiWeb was the first wiki developed in 1995 by Cunningham and there are more than 200 clones today.

The popularity of wikis is due to their simple and flexible nature. Firstly, wikis are characterized by a non-linear hypertext structure: a wiki enables the creation of associative hypertexts where each page contains links to other pages and the user can choose which links to follow. Secondly, wikis are characterized by ease of access since all a user needs to use a wiki is a web browser. As a web application the wiki does not require any specific browser or operating system software. Thirdly, the technical skills required for participation are minimal as users do not have to learn HTML to create a webpage. Thanks to wiki markup, a simplified version of HTML, and the built-in WYSIWYG editors, the burden for creating and formatting a webpage is significantly reduced. Finally, a wiki system also provides other functions such as (a) the tracking of edits through a logging and notification system, (b) the comparison between different versions of edits, (c) rollbacks to earlier versions, (d) discussions per wiki page and (e) customizable read and edit access rights as well as protected pages.

B. Wikis in Education

Wikis have attracted considerable interest in higher education as they can be used to support a wide range of activities including but not limited to documentation, reporting, project management, online glossaries and dictionaries, discussion groups, and information systems. Wikis are particularly suited for advancing learning, collaboration, communication, interaction, sharing, meaning making, and reflection. From an educational point of view, [8] outline three main types of uses for wiki: (a) distributing information: the wiki becomes the medium to disseminate information as well as the place to store the information, (b) collaborative artifact creation: text is the main artifact produced, be it in the form of glossary or other formats, and (c) discussion and review activities: anchored discussion, project case libraries, and peer review.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_81,

While the majority of wiki systems have not been explicitly designed for educational purposes, they can be effectively used to support learning activities in classrooms [9]. Unlike other software tools, the open and flat structure of a wiki means that there are no scaffolds for structuring e.g. the processes of collaboration [8].

Wikis and other Web 2.0 technologies are increasingly being used in higher education. Wikis have been used for supporting various types of educational activities online ([8], [10], [9]). More specifically, wikis have been used for supporting collective learning ([11], [12]), enhancing student interaction [9], online teaching and assessment [13], and reflection [14].

Even though as a rule most of the empirical research on wikis involves small-scale case and exploratory studies, the experiences are, generally speaking, positive (see e.g. [15], [16], [17], [18]). Nevertheless, challenges abound: student resistance ([15], [19]), technical problems ([20], [19]), low level of edits ([21], [22]), and preference for individual work [16].

To date, there are only a few studies addressing either teaching and learning issues or broader pedagogical issues. In fact, as the multiple challenged mentioned above suggest, there is a knowledge gap regarding pedagogies which educators in higher education can employ to effectively promote more student-centered pedagogies. What the evidence from existing research clearly suggests is that a wiki system cannot simply be introduced into a course and effectively foster student collaboration and learning ([15], [19], [18]). The present case study aims to fill this gap by exploring the merit of a pedagogical activity in terms of student collaboration.

C. Rationale

In the study reported here, students used a wiki system for the collaborative creation of a textual artifact. As opposed to being an end in itself, this textual artifact mediated the collaborative development of meaning. Considering that wikis can be used both as tool to create a database reflecting the collective experience and as a repository for this database, students created a database of concepts which reflected their collective understanding. Students collaborated in creating concepts and providing their perspectives on what these concepts meant. Consequently, the collective artifact made represented students' understanding of the concepts involved. More specifically, the students reached an understanding of the course concepts by drawing on (a) lecture slides and notes, (b) course reader, (c) class discussions, and (d) other materials. Students were not simply expected to write concept definitions thereby reiterating information available in the course materials. Instead, the students were expected to write what they thought the concepts meant, how they could be applied to other contexts (e.g. learning and instruction), and how everyday situations could be interpreted using the concepts. The former involved rephrasing the definitions of concepts and providing examples in various contexts. Although this is a first step toward understanding and as such it is essential since students can learn from writing, it requires low level cognitive processing since it involves the reproductive use of existing texts. The latter involves a high level of cognitive processing since it requires thinking about the concept and what it means in relation to other concepts. While students could handle the former without problems by adopting copy and paste strategies, the latter is more challenging because the quest for meaning is an important prerequisite: without meaning it makes no sense.

The process of meaning making is inherently social as meanings are socially constructed, negotiated and interpreted. Wikis are an ideal platform for fostering collective meaning making because they provide opportunities for producing non-linear, complex, elaborate, evolving, multi-author texts. The potential of wikis to contribute to the process of collaborative knowledge creation and sharing has been stressed ([23], [20], [24]).

D. Focus of the study

The present study focuses on the use of a wiki as a means to promote student collaboration and meaning making and examines student collaboration patterns. The study addressed the following questions: (a) *what is the rate of student collaboration in the process of the collaborative creation of a textual artifact?* (b) *what are the most prominent collaboration patterns?*

II. Method

A. Subjects

Fifty students participated in the study. All students had enrolled in an undergraduate course that the author taught in the Department of Preschool Education at the University of Thessaly. The course was compulsory for 5th semester students.

B. Course

The course constituted an introduction to learning theories and the use of ICT to support learning. The course aimed to develop students' knowledge and skills regarding educational software. Regarding knowledge, the course included the familiarization with the various types of educational software as well as the underlying cognitive theories: Behaviorism, Piagetian theory, Cognitive Science, Situated and Distributed Cognition, and Sociocultural Theory. Students were expected to develop a conceptual understanding of a given instructional situation in terms of the learning epistemology adopted. Technology was approached as a cognitive tool and the role it plays in learning was examined. Regarding skills, the course aimed to render students capable for the design, organization, implementation, and evaluation of technology-based instruction. Students were expected to develop the skills to teach a curriculum unit which would be implemented using educational software.

C. Wiki

One of the course assignments involved student participation in a wiki. Participation was compulsory and students received up to 30% of the total course credit. The wiki activity required each student (a) to create five pages related to course concepts and (b) to contribute to five pages on concepts created by other students. More specifically, each student had to select five course concepts and create corresponding wiki pages about those concepts. The wiki pages could refer to any aspect of the concept such as definitions, descriptions, examples, explanations, and applications in teaching and learning contexts. Overall, the students were asked to contribute their own understanding of concepts which had been presented and discussed in class. On the other hand, each student was expected to contribute their understanding to the concepts created by their fellow students in other wiki pages by contributing definitions, descriptions, examples, explanations as well as applications in teaching and learning contexts. As discussed in the rationale section, the idea was that students would advance their own understanding of the concepts involved by contributing to the collective understanding of those concepts as reflected in the wiki pages.

To avoid the low level strategies, the copy and paste of information from other sources was strictly disallowed. Students were instructed to provide references to articles, books or websites whenever they drew on information from such sources. It was emphasized to the students that every idea, interpretation or insight was in principle welcome for it could advance the collective understanding of the group. Student contributed content was routinely scanned for plagiarism by the author and a teaching assistant and on a number of occasions certain students were given notice and asked to rephrase and attribute the origin of the ideas. Finally, it should be noted that no specific writing or stylistic or content instructions were given so no specific genre (e.g. encyclopaedia entry) was privileged over others.

[25], one of the most popular wiki clones, was the server software used for the purposes of the study to host the course wiki. It was installed and specifically customized for the purposes of the course. In the present study the Wiki was used as a tool in a closed study group, i.e. it was only accessible to the students participating in the course. All students who had enrolled in the course had both read and write access to all user-contributed pages. Students were explicitly asked not to lock any pages they had created for that would undermine the very objective of the activity and prevent collaboration. Only a few pages created by the author and the teaching assistant containing instructions and other important information were locked.

D. Procedure

Students were introduced to the wiki at the beginning of the course and its objective was explained. In the sandbox provided students experimented with all technical aspects of the wiki such as creating new pages, revising existing ones, linking, rolling back, and discussing. After each class lecture where several concepts were introduced, analyzed, and discussed, students were asked to select certain concepts which they deemed important and either create wiki pages about them or contribute to the wiki pages made by other students.

In the beginning of the course the students were surveyed regarding background knowledge of and attitudes towards ICT. At the end of the course they were surveyed about their learning experiences with the wiki and were also interviewed using semi-structured interviews.

E. Measures

For the purposes of this study we examined the number of pages created by the students, the number of page views, the number of user views per page, the total number of revisions as well as the number of revisions per page, the number of revisions per user, and the number of user revisions per page. Given the specific research questions, we were particularly interested in (a) the number of revisions per page because it shows how much processing each page underwent, (b) the number of users who made the revisions as it shows how many users were collaboratively involved in revising the page, and (c) the pattern of revisions, namely stylistic/formatting vs. content revisions: the former altered the surface characteristics of the text (such as formatting, layout, style, dividing the text up into smaller units) while the latter addressed the ideas expressed in the text (such as correcting inaccuracies, providing examples, articulating, refining meaning, elaborating, adding definitions, adding more concepts and sub-concepts). The number of page edits as well as the number of unique page editors are typical measures when examining wikis (see e.g. [26], [27]). When it comes to the patterns of revision we followed [28] by distinguishing between content and formatting changes.

III. Analysis

The instructor and teaching assistant contributed pages and views of those pages were discarded from further analysis as they were not informative regarding student collaboration. For the purposes of this study data analysis was mostly exploratory. To address the first research question, data analysis involved frequency counts and indices of central tendency. To address the second research question the student contributions were examined in terms of deep (i.e. content) or surface (formatting) revisions.

IV. Results

A. Extent of student collaboration

Overall, the students created 169 pages in the wiki which were viewed 2902 times with a mean number of student views per page of 17.17. The system recorded a total of 771 revisions, with a mean of 2.53 revisions per page and a mean of 16.06 revisions per user. Finally, the mean number of user revisions per page was 1.85. Based on the data collected, three main findings can be reported.

First, the students did not view all the pages. The total number of views shows less than 3000 page views by students while if all participating students had viewed all pages we would have expected more than 8000 views. The user views per page suggests that on average 17 users viewed the content of any page. This shows that the mean user views per page is low given the number of participating students.

Second, students did not contribute to many pages as judged by the number of revisions per page which is very low. More specifically, the average number of revisions per page is 2.53. This figure is low as it shows that each page has not been subject to many revisions, and the low number of revisions is indicative of low collaboration levels. It should be noted that the number of revisions is an important indicator of collaboration because it shows the level of processing that each page underwent. Even though in principle it might be difficult to define how many revisions would be satisfactory, the overall impression is that this number is rather small. While at first sight this might seem to be promising, the distribution of page views suggests an interesting pattern which can be seen in table 1.

Even though revisions by a single student are important as they provide an indication of the level of processing, work, and elaboration that the student devoted to the task, collaboration is essentially manifested in the joint construction of the textual artifact and more specifically in user revisions of this artifact. As table 1 suggests, the number of pages on which students jointly contributed is very low. More specifically, 40% of the pages did not involve any collaboration whatsoever as the page content was contributed independently by students. In fact the percentage of pages which were created by distinct students is even higher as the examination of pages with 2 revisions indicated that 12 pages were subjected to 2 revisions which were conducted by a single student. As a result, the percentage of pages which did not involve any joint content creation rises to 47%.

Number of pages	% of pages	Number of revisions per page	Mean number of users per revision per page
81	47.93	1	1.00
23	13.61	2	1.66
28	16.57	3	2.14
15	8.88	4	3.20
8	4.73	5	3.63
6	3.55	6	3.67
3	1.78	7	4.00
2	1.18	8	4.00
1	0.59	9	4.00
0	0.00	10	0.00
1	0.59	11	1.00
1	0.59	12	2.00

Table 1: Main wiki statistics

Third, the average number of users who revised a page is 1.85 which is fairly low. The number of users who participated in revising a page is a genuine indicator of collaborative content creation and generally speaking of collaborative meaning making and learning. In this case it appears that there was more than one student involved with contributing content to pages. As can be seen in table 1, about half of the pages were created without any user collaboration. More specifically, about three quarters of all pages were made with little collaboration among users. Thus, in the majority of the wiki pages created, on average one or two students were involved.

B. Patterns of student collaboration

For the purposes of this paper we focused on the 30 pages with the highest number of revisions per page and examined the revision patterns by coding every revision made per page in two categories: (a) surface (b) content. While as mentioned above about half of the wiki pages created did not involve any collaboration (i.e. joint revision) the collaboration pattern which emerged for the other half of the pages is one of addition. More specifically, the analysis revealed a pattern of autonomous additions to the text contributed by others. The examination of revisions indicated that, as a rule, big chunks of text (usually in the form of a short paragraph) were added to existing text. While this form of contribution is important, it signifies a rather low level of revisions and, subsequently, collaboration. Thus, even in the cases of collaboration (i.e. joint creation and revision), the students adopted a cumulative approach: new content was created without altering any of the existing content. This is clearly illustrated in figure 1 where the right column indicates a pattern of successive text additions over time.

Fig 1. Snapshot of the edits of a page

Whenever students collaborated on a page (as indicated by jointly editing it) there were very few edits of stylistic/formatting nature. As a rule, the students did not interfere with the text pages created by others and whenever they did their contributions were cumulative. Interestingly enough, there were no content revisions or elaborations of any kind.

The overwhelming majority of student-created content was based on the lecture slides and notes, course reader, and class discussions. It appears that students rarely referred to other sources and materials.

The examination of the discussions which were linked to each wiki page created indicated that the students did not engage in any discussions with other students regarding the meaning of the concepts. The only discussions recorded by the system were instructor-related ones: questions were posed to the students in an attempt to either emphasize a specific aspect of a concept or to foster student thinking and promote further reflection.

V. Discussion & Conclusions

Given the affordances of Web 2.0 technologies, the present paper explored aspects of a wiki integration into undergraduate teaching. More specifically, the study reported focused on the extent of student collaboration in the collaborative creation of a textual artifact and the most prominent patterns of such collaboration. Collaboration was operationalized as the joint creation of the textual artifact as reflected in the revisions. We found that there were few instances of collaboration as the students did not build on the contributions of others by revising them. This is manifested in the mean number of revisions per page which is fairly low (2.53). In this specific context, collaboration requires building on and revising the work of others and involves both stylistic contributions (e.g. reformatting) and conceptual ones (e.g. changing the content). This finding is in line with the results reported in the literature. For example, [21] found that 15% of the students did not make any edits at all while 25% of the students accounted for over 64% of all edits made by all participants. [29] also reported that most students made few edits while a small number of students edited several pages. [22] concluded that on average every wiki page made was edited twice.

On the one hand, the evidence indicated that about half of the pages created did not involve any joint collaboration as single students were involved in creating the content of the page. It is clear that this pattern of individual contributions is not very promising for collaboration. On the other hand, as far as the other half of the pages are concerned, the analysis indicated a low level approach for revisions: the pattern of contributions was cumulative as students simply added text to the existing text. There was no evidence of extensive elaborations or revisions of any kind. Thus, the picture which emerged from the data suggests that even though the collective creation of a textual artifact was to promote collaboration, the students largely approached it as an individual activity as opposed to a collective one. Drawing conclusions based on this patterns might be somewhat arbitrary given the lack of research on this topic. One can only turn to studies examining the revision patterns in the largest wiki in existence, Wikipedia. It appears that the patterns obtained in this study is rather typical of wikis. For example, [28] reports that additions are the most common revision pattern in Wikipedia. Moreover, [27] concluded that Wikipedia articles accrete over time. Finally, [26] found that the vast majority of Wikipedia edits are additions or deletions with very few structural changes. While this pattern of revisions for Wikipedia appears to hold as it is well documented it should be stressed that it is functional given the objectives and the production model of Wikipedia. For such a context, the observed pattern of revision might be acceptable or of secondary importance. However, in the context of the present study this pattern of edits is indicative of low levels of collaboration and as such it is highly problematic.

As it was discussed in the introduction, even though wikis are becoming widely used in education, the issue of optimal pedagogies remains largely unexplored. The activity which was designed for this study (create five pages on course concepts and contribute to five other pages created by others) turned out to be suboptimal. As the findings suggest, students who are novice wiki users are not very likely to engage in joint revisions and, consequently, collaboration. This shows that wiki systems cannot be readily integrated into undergraduate courses to promote more constructive learning practices. Hence, new tasks, activities, and practices should be devised. As the study findings suggest, from a pedagogical point of view the activity needs to be redesigned so as to promote more interaction among students in the form of collaborative contributions and revisions.

References

[1] Clark, R.C. & Mayer, R.E. (2008). *E-Learning and the Science of Instruction. Proven Guidelines for Consumers and Designers of Multimedia Learning* (2_{nd} edition). San Fransisco: Pfeiffer.

[2] Bonk, C.J., Wisher, R.A. & Lee, J-Y. (2004). Moderating Learner-Centered E-Learning: Problems and Solutions, Benefits and Implications. In T.S. Roberts (Ed). *Online Collaborative Learning: Theory and Practice* (pp. 54-85). London: Idea Group Publishing.

[3] Naidu, S. (2003). Designing Instruction for e-Learning Environments. In M. G. Moore & W.G. Anderson (Eds). *Handbook of Distance Education* (pp. 349-365). Mahwah, New Jersey: LEA.

[4] Conole, G. Oliver, M., Falconer, I., Littlejohn, A. & Harvey, J. (2007). Designing for learning. In G. Conole & M. Oliver (Eds). *Contemporary perspectives in e-learning research: themes, methods and impact on practice* (pp. 101-120). London: Routledge.

[5] Littlejohn, A., Cook, J., Campbell, L., Sclater, N., Currier, S. & Davis, H. (2007). Managing educational rescourses. In G. Conole & M. Oliver (Eds). *Contemporary perspectives in e-learning research: themes, methods and impact on practice* (pp. 134-146). London: Routledge.

[6] Albors, J. Ramos, J.C. & Hervas, J.L. (2008). New learning network paradigms: Communities of objectives, crowdsourcing, wikis and open source. *International Journal of Information Management*, 28, 194–202.

[7] Leuf, B., & Cunningham, W. (2001). *The Wiki Way*. Addison-Wesley Professional.

[8] Guzdial, M. Rick, J. & Kehoe, C. (2001). Beyond Adoption to Invention: Teacher-Created Collaborative Activities in Higher Education. *The Journal of the Learning Sciences*, 10(3), 265–279.

[9] Lund, A. & Smördal, O. (2006). Is there a space for the teacher in a wiki? In WikiSym '06: Proceedings of the 2006 international symposium on Wikis (pp. 37–46). New York: ACM Press.

[10] Augar, N., Raitman, R. & Zhou, W. (2004). Teaching and learning online with wikis. In R. Atkinson, C. McBeath, D. Jonas-Dwyer & R. Phillips (Eds.), *Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference* (pp. 95-104). Perth, 5-8 December. http://www.ascilite.org.au/conferences/perth04/procs/augar.html

[11] Nicol, D., Littlejohn, A. & Grierson, H. (2005). The importance of structuring information and resources within shared workspaces during collaborative design learning. *Open Learning*, 20(1), 31–49.

[12] Forte, A. & Bruckman, A. (2006). From wikipedia to the classroom: exploring online publication and learning. In *ICLS '06: Proceedings of the 7th international conference on Learning sciences*, (pp. 182–188). International Society of the Learning Sciences.

[13] Bruns, A. & Humphreys, S. (2005). Wikis in teaching and assessment: The M/Cyclopedia project. *Proceedings of the ACM WikiSym 05: International Symposium on Wikis*, San Diego, CA, 25-32.

[14] Ras, E. Carbon, R. Decker, B. & Rech, J. (2007). Experience Management Wikis for Reflective Practice in Software Capstone Projects. *IEEE Transactions On Education*, 50(4), 312-320.

[15] Rick, J. & Guzdial, M. (2006). Situating CoWeb: a scholarship of application. *Computer-Supported Collaborative Learning* 1: 89–115.

[16] Elgort, I., Smith, A.G. & Toland, J. (2008). Is wiki an effective platform for group course work? *Australasian Journal of Educational Technology*, 24(2), 195-210.

[17] Robertson, I. (2008). Learners' attitudes to wiki technology in problem based, blended learning for vocational teacher education. *Australasian Journal of Educational Technology*, 24(4), 425-441.

[18] Wheeler, S. Yeomans, P. & Wheeler, D. (2008). The good, the bad and the wiki: Evaluating student-generated content for collaborative learning. *British Journal of Educational Technology*, 39(6), 987–995.

[19] Cole, M. (2008). Using Wiki technology to support student engagement: Lessons from the trenches. *Computers & Education* (2008), doi:10.1016/j.compedu.2008.07.003.

[20] Minocha, S. & Thomas, P. G. (2007). Collaborative Learning in a Wiki Environment: Experiences from a software engineering course. *New Review of Hypermedia and Multimedia*, 13(2), 187-209.
[21] Carr, T. Morrison, A., Cox, G., & Deacon, A. (2007). Weathering wikis: Net-based learning meets political science in a South African university. *Computers and Composition*, 24, 266–284
[22] Ma, W.W-K. & Yuen, A.H-K. (2007). Learning News Writing Using Emergent Collaborative Writing Technology Wiki. In J. Fong, F. L. Wang (Eds). *Blended Learning*. Workshop on Blended Learning (pp. 303-314): Edinburgh, United Kingdom: Pearson.
[23] Raman, M., Ryan, T., Olfman, L. (2005). Designing Knowledge Management Systems for Teaching and Learning with Wild Technology. *Journal of Information Systems Education*, 16(3), 311-320.
[24] Cress, U. & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *Computer-Supported Collaborative Learning*, 3, 105–122.
[25] MediaWiki (2008). http://www.mediawiki.org
[26] Viégas, F. B., Wattenberg, M., & Dave, K. (2004). Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, Vienna, Austria.
[27] Wilkinson, D.M. & Huberman, B.A. (2007). Assessing the value of cooperation in Wikipedia. *First Monday*, 12(4). [On-line: http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1763/1643]
[28] Jones, J. (2008). Patterns of Revision in Online Writing A Study of Wikipedia's Featured Articles. *Written Communication*, 25(2), 262-289.
[29] Ravid, G., Kalman, Y.M. & Rafaeli, S. (2008). Wikibooks in higher education: Empowerment through online distributed collaboration. *Computers in Human Behavior*, 24, 1913–1928.

Actor-Critic Algorithms for Variance Minimization

Yogesh P. Awate {yogesh.awate@iitb.ac.in}
MarketRx - A Cognizant Company, Gurgaon-122002, India[1]

***Abstract*-We consider the framework of a set of recently proposed two-timescale actor-critic algorithms for reinforcement-learning (RL) using the long-run average-reward criterion and linear feature-based value-function approximation. The actor and critic updates are based on stochastic policy-gradient ascent and temporal-difference algorithms, respectively. Unlike conventional RL algorithms, policy-gradient-based algorithms guarantee convergence even with value-function approximation but suffer due to high variance of the policy-gradient estimator. To minimize this variance for an existing algorithm, we derive a stochastic-gradient-based novel critic update. We propose a novel baseline structure for variance minimization of an estimator and derive an optimal baseline which makes the covariance matrix a zero matrix – the best achievable. We derive a novel actor update based on the optimal baseline deduced for an existing algorithm. We derive another novel actor update using the optimal baseline for an unbiased policy-gradient estimator which we deduce from the Policy-Gradient Theorem with Function Approximation. We obtain a novel variance-minimization-based interpretation for an existing algorithm. The computational results demonstrate that the proposed algorithms outperform the state-of-the-art on Garnet problems.**

I. Introduction

Solving Markov decision processes (MDPs) involving large state spaces (curse of dimensionality) necessitates value-function approximation using a parameter vector whose dimensionality is much smaller than the cardinality of the state space. However, conventional methods like policy iteration, value iteration, Q-learning and temporal-difference learning, which converge well with lookup tables, may fail to converge with function approximation [1]-[2]. Policy-gradient methods provide a way out of this. Policy iteration using policy-gradient-based policy improvement step can be proved to be convergent even with function approximation [3]. But this convergence is often slow due to the *high variance* of the policy- gradient estimates obtained through simulation [4]-[6]. Actor-critic algorithms are analogs of policy iteration, with the actor representing the policy and the critic representing the value function. Actor-critic algorithms provide a way for achieving strong convergence guarantees by using a policy-gradient-based actor learning rule and also addressing the problem of *variance reduction* of the policy-gradient estimates using the critic [5]. Bootstrapping critic learning rules such as temporal-difference learning, which decrease variance at the cost of bias, have been popularly used for variance reduction [7, 8]. Policy gradients have also been recently replaced by natural gradients [7]-[10] which were originally discovered by Amari [11]. Natural gradients are covariant [9, 11], lead to faster convergence by taking a "more direct path to the optimal solution in the parameter space" [10] and provide Fisher-efficient online learning [12].

In this paper, Section 2 describes the reinforcement learning (RL) framework which we use. In Section 3, we derive a novel critic update based on variance reduction of the policy-gradient estimator. In Section 4, we introduce a novel baseline structure and derive the optimal baseline, which leads to a novel actor update and an alternative interpretation for one of the existing natural-gradient-based algorithms. In Section 5, we propose another novel actor update using an unbiased policy-gradient estimator which we obtain based on the basic Policy-Gradient Theorem with Function Approximation [3]. In section 6, we study the performance of the proposed algorithms on Garnet problems and make some interesting observations.

II. Reinforcement Learning Framework

We consider the RL framework as in [7, 8], which is the most sophisticated amongst previously considered RL frameworks e.g. in [1]-[6], [9], [10] and [14]. We have a finite state space *S={1,2,...,n},* finite action space *A* and random uniformly bounded immediate rewards. All actions are possible in each state. Let $s_t \in S$, $a_t \in A$ and $r_t \in \mathbb{R}$ denote the state, action and immediate reward, respectively, at time $t \in \{0,1,2,...\}$. Let $P(s,a,s') = \Pr(s_{t+1} = s' \mid s_t = s, a_t = a\}$ denote the one-step transition probabilities and let $R(s,a) = E(r_{t+1} \mid s_t = s, a_t = a\}$ denote the single-stage expected rewards. A parameterized randomized stationary policy π^θ : $\theta \in R^{d_1}$ maps each state *s* to a probability distribution $\pi^\theta(s,.)$ over *A*. Let $J(\pi^\theta)$ or equivalently $J(\theta)$ denote the long-run average-reward under the policy π^θ and $\{s_t : t = 0,1,2,...\}$ denote the MDP.

Assumption (A1) [7, 8]: "Under any policy π, the Markov chain resulting from the MDP $\{s_t : t = 0,1,2,...\}$ is irreducible and aperiodic."

Let the unique stationary probability distribution be $d^\pi = \{d^\pi(s), s \in S\}$. In the rest of the paper, we use the short-

[1] This work was done when the author was a graduate student at the Department of Industrial Engineering and Operations Research, Indian Institute of Technology, Bombay.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_82, © Springer Science+Business Media B.V. 2010

hand notation $E_{(s,a)}$ for $E_{s\sim d^{\pi}(.), a\sim\pi(s,.)}$. The long-run average-reward is given by:

$$J(\pi) = \lim_{T\to\infty}\frac{1}{T}E\left[\sum_{t=0}^{T-1} r_{t+1} \mid \pi\right] = \sum_{s\in S} d^{\pi}(s)\sum_{a\in A}\pi(s,a)R(s,a)$$

Our objective is to find the optimal policy $\theta^{opt} = \arg\max_{\theta} J(\theta)$

The action-value function for the state-action pair (s,a) under policy π is defined as:

$$Q^{\pi}(s,a) = \sum_{t=0}^{\infty} E\left[r_{t+1} - J(\pi) \mid s_0 = s, a_0 = a, \pi\right]$$

As dictated by the Policy-Gradient Theorem with Function Approximation [3], a linear feature-based approximation is used as $\hat{Q}^{\pi}_{w^*}(s,a) = w^{*T}\psi_{sa}$ where w^* is a stationary point of the MSE

$$\mathcal{E}^{\pi}(w) = \sum_{s\in S} d^{\pi}(s)\sum_{a\in A}\pi(s,a)\left[Q^{\pi}(s,a) - \hat{Q}^{\pi}_{w}(s,a)\right]^2$$

$$\text{i.e. } \frac{\partial}{\partial w}\mathcal{E}^{\pi}(w)\bigg|_{w=w^*} = 0 \text{ and } \psi_{sa} = \nabla_{\theta}\log\pi(s,a)$$

are called *compatible* state-action feature vectors. It is known that w^* is also the natural-gradient of the average-reward [9]. The state-value function for the state s under policy π is given by $V^{\pi}(s) = \sum_{a\in A}\pi(s,a)Q^{\pi}(s,a)$. A linear feature-based approximation for the state-value function is used as $\hat{V}^{\pi}(s) = v_t^T f_s$ where $v_t \in R^{d_2}$ is the critic weight-vector at time *t* and $f_s \in R^{d_2}$ is the feature-vector for state *s*. The algorithms are two-timescale [13] with the actor and critic on the slower and faster timescales characterized by the step size sequences (β_t) and (α_t), respectively. Figure 1 shows the structure of two-timescale actor-critic algorithms that we consider (see [7, 8]).

1: **Input**: $\pi^{\theta}(.,.), f_S$
2: **Initialize:** $\theta = \theta_0, v = v_0, \alpha = \alpha_0, \beta = \beta_0,$
$\xi = c\alpha_0, \hat{J}_0 = 0, s_0$ (randomly)
3: **for** *t=0; ! stopping condition*; *t=t+1*
3a: Select action $a_t \sim \pi^{\theta_t}(s_t,.)$
3b: Observe next state $s_{t+1} \sim P(s_t, a_t, .)$ and immediate reward r_{t+1}
3c: Maintain a step-size schedule satisfying the conditions:

$$\sum_t \alpha_t = \sum_t \beta_t = \infty,\ \sum_t\left(\alpha_t^{\,2} + \beta_t^{\,2}\right) < \infty,$$
$$\beta_t = o(\alpha_t),\ \xi_t = c\alpha_t$$

3d: Perform updates:
3d: (i) **Average reward:**

$$\hat{J}_{t+1} = (1-\xi_t)\hat{J}_t + \xi_t r_{t+1}$$

3d: (ii) **TD error:**

$$\delta_t = r_{t+1} - \hat{J}_{t+1} + v_t^T f_{s_{t+1}} - v_t^T f_{s_t}$$

3d: (iii) **Critic**: algorithm specific
3d: (iv) A**ctor:** algorithm specific
4: **Return:** policy parameter vector θ, value function weight vector v

III. A Novel Critic Update For Variance Reduction Of The Policy-Gradient Estimator

The analysis in this section assumes the framework of Algorithm 1 of [7, 8].

Algorithm 1 [7, 8]:

Critic Update: $v_{t+1} = v_t + \alpha_t\delta_t f_{s_t}$ (1)

Actor Update: $\theta_{t+1} = \theta_t + \beta_t\delta_t\psi_{s_t a_t}$ (2)

In this algorithm, $\delta_t^{\pi}\psi_{s_t a_t}$ is used as a biased estimator for the average-reward gradient $\nabla_{\theta}J(\pi)$ (Lemma 4 of [7, 8]), where

$$\lim_{t\to\infty} v_t = v^{\pi}\ w.p.1,\ \delta_t^{\pi} = r_{t+1} - \hat{J}_{t+1} + v^{\pi^T} f_{s_{t+1}} - v^{\pi^T} f_{s_t}.$$

In the actor update, $\delta_t \approx \delta_t^{\pi}$, as asymptotically the slower actor sees the faster critic as nearly equilibrated [13]. The covariance matrix for $\delta_t^{\pi}\psi_{s_t a_t}$ is given by:

$$\operatorname{cov}\left(\delta_t^{\pi}\psi_{s_t a_t}\right) = E_{(s,a)}\left[\delta_t^{\pi}\psi_{s_t a_t} - E_{(s,a)}\left(\delta_t^{\pi}\psi_{s_t a_t}\right)\right]\left[\delta_t^{\pi}\psi_{s_t a_t} - E_{(s,a)}\left(\delta_t^{\pi}\psi_{s_t a_t}\right)\right]^T \quad (3)$$

We seek to minimize the trace of this covariance matrix given by:

$$Trace\left(\operatorname{cov}\left(\delta_t^{\pi}\psi_{s_t a_t}\right)\right) = E_{(s,a)}\left[\delta_t^{\pi}\psi_{s_t a_t} - E_{(s,a)}\left(\delta_t^{\pi}\psi_{s_t a_t}\right)\right]^T\left[\delta_t^{\pi}\psi_{s_t a_t} - E_{(s,a)}\left(\delta_t^{\pi}\psi_{s_t a_t}\right)\right] \quad (4)$$

Following a stochastic gradient descent approach, we use

$$v_{t+1} - v_t = -\alpha_t\nabla_{v_t}\left[\delta_t^{\pi}\psi_{s_t a_t} - E_{(s,a)}\left(\delta_t^{\pi}\psi_{s_t a_t}\right)\right]^T\left[\delta_t^{\pi}\psi_{s_t a_t} - E_{(s,a)}\left(\delta_t^{\pi}\psi_{s_t a_t}\right)\right] \quad (5)$$

This leads to the desired critic update as

$$v_{t+1} - v_t = -\alpha_t\left[\nabla_{v_t}\left[\delta_t^{\pi}\psi_{s_t a_t} - E_{(s,a)}\left(\delta_t^{\pi}\psi_{s_t a_t}\right)\right]\right]\left[\delta_t^{\pi}\psi_{s_t a_t} - E_{(s,a)}\left(\delta_t^{\pi}\psi_{s_t a_t}\right)\right] \quad (6)$$

The RHS is a function of v^π and the current estimate of v^π is v_t. Hence, replacing v^π by v_t (and hence, δ_t^π by δ_t) throughout, we derive and obtain:

$$\nabla_{v_t}\left(\delta_t\psi_{s_t a_t} - E_{(s,a)}\left(\delta_t\psi_{s_t a_t}\right)\right) = \left(f_{s_{t+1}} - f_{s_t}\right)\psi_{s_t a_t}{}^{\mathrm{T}} - \sum_{s\in S} d^\pi(s)\sum_{a\in A}\pi(s,a)\sum_{s'\in S}P(s,a,s')f_{s'}\nabla_\theta \log\pi(s,a)^T \quad (7)$$

Replacing the expectation by a single estimate, we get

$$\nabla_{v_t}\left[\delta_t\psi_{s_t a_t} - E_{(s,a)}\left(\delta_t\psi_{s_t a_t}\right)\right] = \left(f_{s_{t+1}} - f_{s_t}\right)\psi_{s_t a_t}{}^{\mathrm{T}} - f_{s_{t+1}}\psi_{s_t a_t}{}^{\mathrm{T}} = -f_{s_t}\psi_{s_t a_t}{}^{\mathrm{T}} \quad (8)$$

Hence, (6) becomes:

$$v_{t+1} - v_t = \alpha_t f_{s_t}\psi_{s_t a_t}{}^{\mathrm{T}}\left[\delta_t\psi_{s_t a_t} - E_{(s,a)}\left(\delta_t\psi_{s_t a_t}\right)\right] \quad (9)$$

We use an exponentially smoothed average τ_{t+1} to approximate $E_{(s,a)}\left(\delta_t\psi_{s_t a_t}\right)$ as follows:

$$E_{(s,a)}\left[\delta_t\psi_{s_t a_t}\right] \approx \tau_{t+1} = \alpha_t\delta_t\psi_{s_t a_t} + (1-\alpha_t)\tau_t,\ \tau_0 = 0 \quad (10)$$

$E_{(s,a)}\left(\delta_t\psi_{s_t a_t}\right)$ is a function of θ_t and, hence, τ_{t+1} is updated on the faster timescale characterized by (α_t). Hence, (9) becomes

$$v_{t+1} - v_t = \alpha_t f_{s_t}\psi_{s_t a_t}{}^{\mathrm{T}}\left[\delta_t\psi_{s_t a_t} - \tau_{t+1}\right] \quad (11)$$

Hence, we have the following **Algorithm (A1)** with the *novel critic update***:**

Exponentially-Smoothed Average Update: $\tau_{t+1} = \alpha_t\delta_t\psi_{s_t a_t} + (1-\alpha_t)\tau_t,\ \tau_0 = 0$
Critic Update: $v_{t+1} = v_t + \alpha_t f_{s_t}\psi_{s_t a_t}{}^{\mathrm{T}}\left(\delta_t\psi_{s_t a_t} - \tau_{t+1}\right)$
Actor Update: $\theta_{t+1} = \theta_t + \beta_t\delta_t\psi_{s_t a_t}$

IV. A Novel Actor Update Based On A Novel BaseLine Structure

We subtract a baseline control variate from the average-reward gradient estimator $\delta_t\psi_{s_t a_t}$, to give a new estimator. To ensure that the subtracted baseline does *not* introduce any additional bias in the average-reward gradient estimation, we consider only the class of baselines with expectation zero. We consider a novel class of baseline functions having the form $\left(X_t - E\left(X_t\right)\right)$, where X_t is a statistic known at time *t*. Hence, the average-reward gradient estimator after using the baseline becomes $\delta_t\psi_{s_t a_t} - \left(X_t - E\left(X_t\right)\right)$ and

$$E\left(\delta_t\psi_{s_t a_t} - \left(X_t - E\left(X_t\right)\right)\right) - \nabla_\theta J(\pi) = E_{(s,a)}\left(\delta_t\psi_{s_t a_t}\right) - \nabla_\theta J(\pi) \quad (12)$$

Within this class of baseline functions, we call the baseline which minimizes $Trace\left(\mathrm{cov}\left(\delta_t\psi_{s_t a_t} - \left(X_t - E\left(X_t\right)\right)\right)\right)$ as the optimal baseline. Upon careful observation, it is clear that the optimal baseline is $\delta_t\psi_{s_t a_t} - E_{(s,a)}\left(\delta_t\psi_{s_t a_t}\right)$ corresponding to $X_t = \delta_t\psi_{s_t a_t}$. This optimal baseline, in fact, makes the matrix $\mathrm{cov}\left(\delta_t\psi_{s_t a_t} - \left[\delta_t\psi_{s_t a_t} - E_{(s,a)}\left(\delta_t\psi_{s_t a_t}\right)\right]\right)$ zero. Making the covariance matrix of the estimator a zero matrix is the best any baseline can achieve. In general, this implies that the optimal baseline for an estimator $\hat\theta$ is $\hat\theta - E(\hat\theta)$. The optimal-baseline approximation that we use is $\delta_t\psi_{s_t a_t} - \tau_{t+1}$, where τ_{t+1} is defined and updated as per (10). Hence, the novel actor update is:

$$\theta_{t+1} = \theta_t + \beta_t\left[\delta_t\psi_{s_t a_t} - \left(\delta_t\psi_{s_t a_t} - \tau_{t+1}\right)\right] = \theta_t + \beta_t\tau_{t+1} \quad (13)$$

This leads to the following **Algorithm (A2)** with the *novel actor update*:

Critic Update: $v_{t+1} = v_t + \alpha_t\delta_t f_{s_t}$
Exponentially-Smoothed Average Update: $\tau_{t+1} = \alpha_t\delta_t\psi_{s_t a_t} + (1-\alpha_t)\tau_t,\ \tau_0 = 0$
Actor Update: $\theta_{t+1} = \theta_t + \beta_t\tau_{t+1}$

A. Observations Regarding Previous Baselines

The approach in [7, 8], first, seems *not* to distinguish between mean squared error (MSE) and variance (e.g. in Lemma 2 of [7, 8]). Second, the baseline in [7,8] concerns the estimation of $Q^\pi(s,a)$ by $\hat Q_w^\pi(s,a) = w^T\psi_{sa}$ and *not* the estimation of the average-reward gradient. Thus, the optimal baseline considered in [7, 8] is *not for variance minimization* of the average-reward gradient estimator, but for the MSE minimization in the estimation of $Q^\pi(s,a)$ by $\hat Q_w^\pi(s,a) = w^T\psi_{sa}$.

The approach in [6] does *not* consider the *covariance matrix structure* of the policy-gradient estimator and seems to treat policy-gradient vectors as scalars. Thus, the derivation of the optimal baseline in Theorem 8 in [6] is, in effect, *restricted* to only a *one-dimensional* policy parameter vector.

B. A New Interpretation of Algorithm 4 [7, 8] Through an Optimal Baseline Framework

In Algorithm 2 of [7, 8], $G_t^{-1}\delta_t\psi_{s_t a_t}$ is used as an estimator of the natural-gradient. The natural-gradient estimator, after incorporating the optimal baseline structure described in Section 4, becomes $E_{(s,a)}\left(G_t^{-1}\delta_t\psi_{s_t a_t}\right)$ and can be approximated as:

$$E_{(s,a)}\left[G_t^{-1}\delta_t\psi_{s_t a_t}\right] \approx \kappa_{t+1} = \alpha_t G_t^{-1}\delta_t\psi_{s_t a_t} + (1-\alpha_t)\kappa_t,\ \kappa_0 = 0 \quad (14)$$

We, thus, have the novel natural-gradient-based actor update as:

$$\theta_{t+1} = \theta_t + \beta_t \kappa_{t+1} \tag{15}$$

Note that (15) is identical with the actor update of Algorithm 4 of [7, 8] which uses w_{t+1} as an estimate of the natural-gradient. w_{t+1} and κ_{t+1} are in fact proxies of each other. Thus, the baseline approach derived here provides a new optimality-based interpretation of Algorithm 4 of [7, 8] in terms of variance minimization of the natural-gradient estimator.

V. A Novel Policy-Gradient Based Algorithm Using The Policy-Gradient Theorem With Function Approximation

From the Policy-Gradient Theorem with Function Approximation [3], we have:

$$\nabla_\theta J(\pi) = \sum_{s\in S} d^\pi(s) \sum_{a\in A} \nabla_\theta \pi(s,a) \hat{Q}^\pi_{w^*}(s,a) \tag{16}$$

where w^*, $\hat{Q}^\pi_{w^*}(s,a)$ are as defined in Section 2. Thus, we have

$$\nabla_\theta J(\pi) = E_{s\sim d^\pi(.), a\sim\pi(s,.)}\left[\psi_{sa}\psi_{sa}{}^T w^*\right] \tag{17}$$

Hence, $\psi_{s_t a_t}\psi_{s_t a_t}{}^T w^*$ is an unbiased estimator of the average-reward gradient. The average-reward gradient estimator after incorporating the optimal baseline becomes

$$\psi_{s_t a_t}\psi_{s_t a_t}{}^T w^* - \left[\psi_{s_t a_t}\psi_{s_t a_t}{}^T w^* - \underset{(s,a)}{E}\left(\psi_{s_t a_t}\psi_{s_t a_t}{}^T\right) w^*\right]$$
$$= \underset{(s,a)}{E}\left(\psi_{s_t a_t}\psi_{s_t a_t}{}^T\right) w^* \tag{18}$$

An estimate of $w^*(\theta_t)$ can be obtained for use in the actor update by the stochastic gradient descent rule i.e. by updating its estimate w_t on the faster timescale along the negative instantaneous gradient direction of the MSE $\mathcal{E}^{\theta_t}(w_t)$, which gives the update rule (also used in Algorithm 3 in [7, 8]) as

$$w_{t+1} = \left[I - \alpha_t \psi_{s_t a_t}\psi^T_{s_t a_t}\right] w_t + \alpha_t \delta_t \psi_{s_t a_t},\ w_0 = 0 \tag{19}$$

Also, $\underset{(s,a)}{E}\left(\psi_{s_t a_t}\psi_{s_t a_t}{}^T\right)$ is the Fisher matrix and can be approximated using an exponentially-smoothed average (as in Algorithm 2 of [7, 8]) as:

$$\underset{(s,a)}{E}\left(\psi_{s_t a_t}\psi_{s_t a_t}{}^T\right) \approx G_{t+1} = (1-\alpha_t)G_t + \alpha_t \psi_{s_t a_t}\psi_{s_t a_t}{}^T,$$
$$\textit{with } G_0 = k[I]_{d_1\times d_1},\ k>0 \tag{20}$$

Hence, the average-reward gradient estimator $\underset{(s,a)}{E}\left(\psi_{s_t a_t}\psi_{s_t a_t}{}^T\right) w^*$ can be approximated as

$$\underset{(s,a)}{E}\left(\psi_{s_t a_t}\psi_{s_t a_t}{}^T\right) w^* \approx G_{t+1} w_{t+1} \tag{21}$$

This leads to the following *novel* **Algorithm (A3)**:

Critic Updates: $v_{t+1} = v_t + \alpha_t \delta_t f_{s_t}$

$$w_{t+1} = \left[I - \alpha_t \psi_{s_t a_t}\psi^T_{s_t a_t}\right] w_t + \alpha_t \delta_t \psi_{s_t a_t}$$

Fisher-matrix estimate update:

$$G_{t+1} = (1-\alpha_t)G_t + \alpha_t \psi_{s_t a_t}\psi_{s_t a_t}{}^T$$

Actor update: $\theta_{t+1} = \theta_t + \beta_t G_{t+1} w_{t+1}$

VI. Results

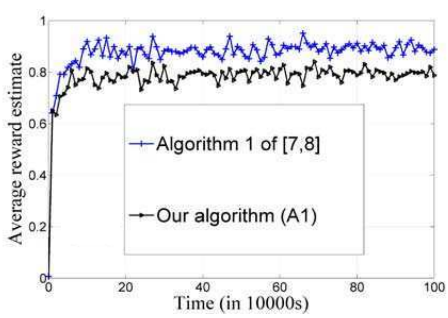
For Garnet (generic average-reward non-stationary environment testbed) specification, we use the notation (n, m, b, σ, d, l) to denote, respectively, (number of states, number of actions, branching factor, standard deviation of the randomly generated immediate reward, dimensionality of the state feature-vector, number of ones in the state feature vector). Since we construct the Garnet instances exactly as specified in [7], we omit further details here. The computational results in [7] indicate superior performance of algorithms in [7, 8] over algorithms in [14], for Garnet problems. Hence we do not show comparisons with algorithms in [14]. As the proposed algorithms are policy-gradient based and the only policy-gradient based algorithm of [7, 8] is Algorithm 1 (others are natural-gradient based), we compare the proposed algorithms with Algorithm 1 of [7, 8]. Also, since results in [7] indicate that their Algorithm 3 performs the best, we also show comparisons with Algorithm 3. For Algorithms 1 and 3 of [7, 8], we use the step-size parameter settings as given in [7]. In Algorithms (A1) and (A2) we used the step-size schedule $(0.1\alpha_t)$ in place of (α_t) for the exponentially-smoothed average update in (10) and we used the same step-size parameter settings otherwise as in Algorithm 1 of [7]. In Algorithm (A3), we use the step-size schedule $(0.001\alpha_t)$ and $k=1$ for the Fisher-matrix update in (20) and used the step-size parameter settings (0.1, 0.01,100000, 1000) for $(\alpha_0, \beta_0, \alpha_c, \beta_c)$, respectively.

From here on, we refer to Algorithms 1 and 3 of [7, 8] as (B1) and (B3), respectively. Figures 2(a1)-(a2), 2(b) and 2(c) are obtained by averaging the average-reward estimates over 20 runs (each run being a different Garnet instance) of 1 million iterations for three sets S_1, S_2 and S_3, respectively; each set consists of 20 randomly generated Garnet (30,4,2,0.1,8,3) instances. Comparing the average of the average-reward estimates, Algorithm (B1) is better than (A1) for the set S_1, as shown in Figure 2(a1). We obtained a definite improvement by a using a convex combination of the critic updates of Algorithms (A1) and (B1). For example, as shown in Figure 2(a2), we found that for the set S_1, a 0.7:0.3 convex combination of the critic updates of Algorithms (A1) and (B1) led to the average of the average-reward estimates being higher than that of (B1), and as high as and more stable than (B3). We don't show here the detailed results for different convex combination parameters, given limited space. For the set S_2, Algorithm (A2) gave the average of the average-reward estimates higher than (B1), as shown in Figure 2(b). For the set S_3, Algorithm

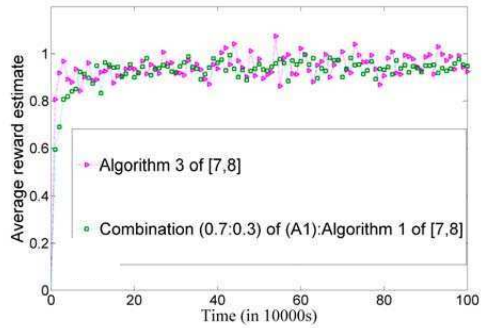
(A3) gave the average of the average-reward estimates higher than (B1), as shown in Figure 2(c).

Through computations on hundreds of Garnet instances, we found that there was no single algorithm which performed the best for all instances. The results in [7] indicate that (B3) gives the highest average-reward estimate amongst all their algorithms. However, Figure 3(a) shows that for a randomly generated instance I_1 of Garnet (30,4,2,0.1,8,3), (B3) gives a lower average-reward estimate than (B1) which in turn is lower than our Algorithm (A2). In Figure 3(a), for Algorithm (A2), we used the step-size schedule $(0.001\alpha_t)$ in place of $(0.1\alpha_t)$ mentioned earlier for the exponentially-smoothed average update in (10). Our Figures 3(b1) and 3(b2) indicate that our Algorithms (A1), (A2) and (A3) have average-reward estimates higher than (B1) for a randomly generated instance I_2 of Garnet (30,4,2,0.1,8,3). Algorithm (A1) looks particularly promising with its average-reward estimate being as high as and much more stable than Algorithm 3 of [7, 8].

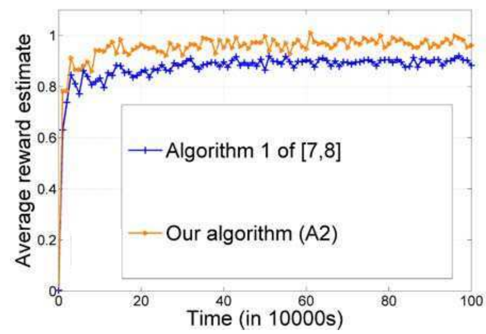
The authors in [7] state that for one of their Garnet (30,4,2,0.1,8,3) instances, the average of the average-reward estimates for each of the algorithms in [7] converged to the unconstrained (without function approximation) optimal average-reward. We found, over experimentation involving hundreds of runs (covering independent runs for different Garnet instances as well as independent runs over the same Garnet instance), that the terminal average-reward estimate nearly matched the unconstrained optimal average-reward hardly for a few percent of the runs. Furthermore, we found that the discrepancy between the terminal average-reward estimate and the unconstrained optimal average-reward varied from a few percent to figures exceeding 50% in many runs for several instances for Algorithm (B3). Our experiments indicate that the loss of optimality due to value-function approximation, is quite prominent for Garnet (30,4,2,0.1,8,3) problems, contrary to what [7] suggests---that the loss is absent for the small Garnet (30,4,2,0.1,8,3) problems.



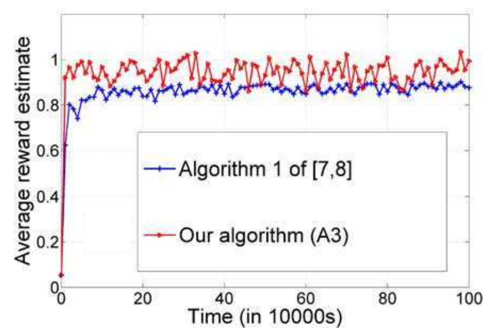
(a1)



(a2)

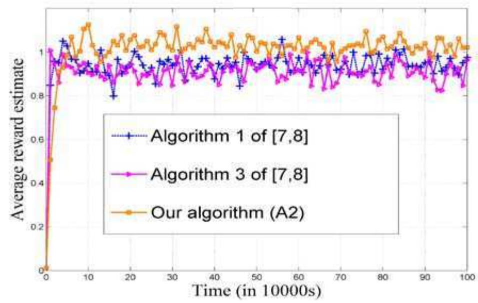


(b)

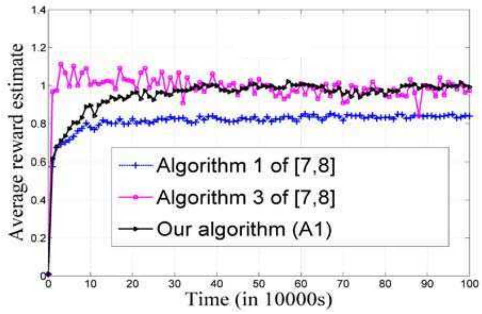


(c)

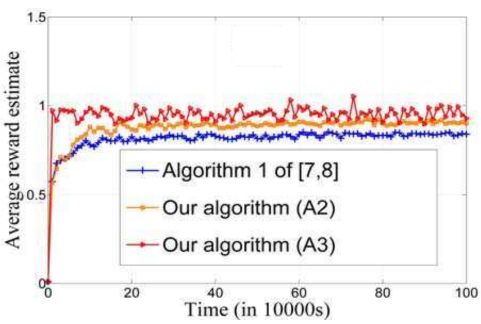
Figure 2. Parts (a1) and (a2) (split for clarity) compare the average of the average-reward estimates over 20 runs (each run being a different Garnet instance) of 1 million iterations of Algorithms 1 and 3 of [7,8] and the proposed novel Algorithm (A1) as well as a convex combination of (A1) and Algorithm 1 of [7,8], on a set S_1 of 20 randomly generated instances of Garnet (30, 4, 2, 0.1, 8, 3). Part (b) gives a similar comparison between Algorithm 1 of [7, 8] and proposed novel Algorithm (A2) for a set S_2 of 20 randomly generated instances of Garnet (30, 4, 2, 0.1, 8, 3). Part (c) compares Algorithm 1 of [7, 8] and proposed novel Algorithm (A3) for a set S_3 of 20 randomly generated instances of Garnet (30, 4, 2, 0.1, 8, 3).



(a)



(b1)



(b2)

Figure 3. Part (a) compares the average of the average-reward estimates over 20 runs of 1 million iterations of Algorithms 1 and 3 of [7,8] and the proposed novel Algorithm (A2) on a single instance I_1 of Garnet (30, 4, 2, 0.1, 8, 3). Parts (b1) and (b2) (split for clarity) compare the average of the average-reward estimates over 40 runs of 1 million iterations, of Algorithms 1 and 3 of [7, 8] and the proposed novel Algorithms (A1), (A2) and (A3), on a single instance I_2 of Garnet (30,4,2,0.1,8,3).

VII. Conclusions

Under the framework of two-timescale actor-critic algorithms with value-function approximation considered in [7, 8], we have derived a *novel critic update in Algorithm (A1)* and *two novel actor updates in Algorithms (A2) and (A3)*. All these intend to minimize the *variance* of the average-reward gradient estimator. The critic update in (A1) is based on the stochastic gradient descent method. For reducing the variance of an estimator without the addition of any bias, we consider a *novel baseline control variate structure*. The actor update in (A2) uses the *optimal baseline* intended for variance minimization of the average-reward gradient estimator of Algorithm 1 of [7, 8]. We *derive* an *unbiased* policy-gradient estimator based on the Policy-Gradient Theorem with Function Approximation [3]. We apply the optimal baseline structure to this estimator to obtain the novel actor update in (A3). We find the optimal baseline derivation in [6] to be valid only for a one-dimensional policy parameter vector. We obtain an *alternative derivation* for Algorithm 4 of [7, 8] based on our optimal baseline structure applied to variance minimization of the natural-gradient estimator used in Algorithm 2 of [7, 8].

Our computational results on randomly generated Garnet problems reveal that the proposed *novel algorithms outperform the state of the art* (algorithms of [7, 8]).

Acknowledgment

The author would like to sincerely thank Prof. N. Hemachandra of the Department of Industrial Engineering and Operations Research, Indian Institute of Technology, Bombay for useful discussions and for providing a preprint of [7].

References

[1] D.P. Bertsekas and J.N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, MA, 1996.

[2] L. Baird, "Residual algorithms: reinforcement learning with function approximation", *Proc. 12th International Conf. on Machine Learning*, 1995, pp 30-37.

[3] R. Sutton, D. McAllester, S. Singh and Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation", *Adv. in Neural Info. Proc. Systems*, 2000, 12:1057-1063.

[4] P. Marbach and J.N. Tsitsiklis, "Simulation-based optimization of Markov reward processes", *IEEE Trans. on Automatic Control*, 2001, 46:191-209.

[5] J. Baxter and P.L. Bartlett, "Infinite-horizon policy-gradient estimation", *Journal of Artificial Intelligence Research*, 2001, 15:319-350.

[6] E. Greensmith, P.L. Bartlett and J. Baxter, "Variance reduction techniques for gradient estimates in reinforcement learning" *Journal of Machine Learning Research*, 2004, Vol. 5, pp. 1471–1530.

[7] S. Bhatnagar, R.S. Sutton, M. Ghavamzadeh and M. Lee, "Natural-gradient actor-critic algorithms" *Automatica, 2007* (to appear, http://drona.csa.iisc.ernet.in/~shalabh/pubs/ac_ bhatnagar.pdf).

[8] S. Bhatnagar, R.S. Sutton, M. Ghavamzadeh and M. Lee, "Incremental natural actor-critic algorithms", *Proc. 21st Annual Conference on Neural Information Processing Systems*, 2007.

[9] S. Kakade, "A natural policy gradient", *Adv. in Neural Info. Proc. Systems*, 2002, 14.

[10] J. Peters, S. Vijayakumar and S. Schaal, "Natural actor-critic", *Proc. 16th European Conference on Machine Learning*, 2005, pp. 280-291.

[11] S. Amari, K. Kurata and H. Nagaoka, "Information geometry of Boltzmann machines" *IEEE Trans. on Neural Networks*, 1992, Vol. 3, No. 2, pp 260-271.

[12] S. Amari, "Natural gradient works efficiently in learning", *Neural Computation*, 1998, 10(2):251-276.

[13] V.S. Borkar, "Stochastic approximation with two timescales", *Systems and Control Letters*, 1997, 29:291-294.

[14] V.R. Konda and J.N. Tsitsiklis, "On actor-critic algorithms", SIAM Journal on Control and Optimization, 2003, 42(4):1143-1166.

e-Education as a State Service to Citizens of Different Ages

Deniss Kumlander
Department of Informatics, TTU
Tallinn, Estonia
e-mail: kumlander@gmail.ee

Abstract—**The paper concentrates on the life-long education. The rapidly changing world is a challenge for both individuals and states as the performance of the work-force is one of the most important factors of any state's long term success in both financial and social sectors. The paper proposes an e-education system that should be funded by the state and exposed to all citizens in order to constantly upgrade their skills. It is a state service targeted to individuals of different ages as an alternative to the massive education targeted exclusively to the young people. The paper also outlines the model of such system and suggests the way it can be built.**

Keywords-e-education; boundary spanning; life-long learning; state services

I. INTRODUCTION

The typical approaches to the educational system in many countries up to now was to organise in the best possible ways universities, colleges and other similar centres massively educating young people. Those centres of education are funded constantly building modern research and study facilities in order to produce highly skilled work-force which is so much required by the modern world [1].

This approach has been criticised a lot [2, 3]. So far universities were mostly looking into the past as served as accumulators of knowledge and were spreading this knowledge during decades without major updates of courses [3]. Fortunately educational centres role is changing at the moment, but is still not ideal. First of all, the university education is too far from technologies been applied in the private sector and is mostly concentrates on dealing with fundamental science and approaches [4]. Universities do argue that this is much more stable method of education as proper understanding of fundamental science (basics, techniques) will give to students a good base for acquiring whatever knowledge in the future [5]. Having agreed with it, we still like to highlight the sufficient gap between the knowledge obtained in universities and the knowledge students will need dealing with specific tasks at their future workplaces. The colleges approaches looks to be providing the required knowledge, but in the modern world this knowledge outdates very quickly and so this type of education has also certain disadvantages. Notice that all earlier mentioned approaches are based on the belief that it is enough to obtain skills once in the beginning of the life.

Unfortunately, this approach to the massive education doesn't work any longer in the modern, rapidly changing world very well. Constant technology, business processes, political and standards changes produce in fact different types of unemployment (for example the structural one), low productivity of workers etc as skills and knowledge become outdated very quickly affecting people capability to do their work effectively. This process affects all involved parties: business, individuals and the entire country as levels of salaries or unemployment define other important aspects of the society like the crime level, peoples believes into society and common, shared set of values that could start to include destructing elements like alcohol or drugs instead of "truth" etc. Moreover peoples low educational, salaries and employment level will result in increasing (and sometimes sufficiently) financing required by different social programs that should be executed to support certain percentage of people that were thrown out from the modern life. All those costs could overweight in total the investment that the state-wide e-educational system could ever demand in order to provide good enough service for people and help them to upgrade skills or obtain a new profession, to make then competitive once again on the market or increase their productivity.

There is another approach suggesting the life-long education, but so far is has been seen either to be up to each individual or to complex and expensive to be supported by the state. Therefore each country government will rather fund universities for massive education of young people building research facilities etc than to invest into local re-education centres. Fortunately the time is changing and the importance of an electronic learning as an alternate solution to the traditional education has sufficiently increased recently with development of electronic channels and world globalization [6, 7]. There are millions of students already attending e-courses in developed countries and this number will increase when more countries will be involved.

The paper does propose to join the need to re-educate the current workforce in order to ensure their high knowledge quality and the modern approaches to the electronic channels based education building up a state-wide educational system funded by the government and available for citizens of different ages.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_83,

II. State Wide Educational System

A. Targets

The re-education process organisation greatly depends on the type of knowledge to be obtained. There are a lot of areas that demands hands-on experience to be obtained in special centres with corresponding equipment. Obviously organisation and financing of such centres is very complex question and is beyond this article scope. Here we would like to discuss such subjects to be studied that do not demand expensive equipment to be used in order to obtain new skills or a new profession. Fortunately, the new era of IT centric technologies and robotic has decreased the need of brutal force employment and increased office type work where skills like languages, computer programming and computers use in different organisational processes are highly demanded. All this can be learned mostly using computer based virtual environment with very small or absence at all of hands-on practise.

B. Auditorium

Sometimes the process of re-education is compared to the process of education (from colleges to universities) and even more often those processes are unified. Such approach has both positive and negative sides. Unification of processes decreases costs and reusability enables using same courses (course materials) for very different environments and groups of people. At the same time we rarely have some kind average person presented widely. Commonly saying the average one is the person, who never existed at all and shows no more than the average temperature for a hospital. The educational process (based on all advantages of unification so far) should be much more selective been applied. It is possible to define that the virtual education environment should be built basing on the context sensitive systems approach were the general system is capable to adapt to the concrete environment and persons it communicates to.

Considering the country wide educational system we need to notice that the system doesn't deal with young students that see the education process rather like a process to pass thrown than a process to learn. They appreciate much more the final result rather than knowledge obtained during the study as see the diploma as a ticket to the happy and rich future. Unlike this category of education process participants, people that would like to re-educate do realise that the diploma gives much less than the real knowledge, which actually does ensure the future workplace and its stability. Besides re-educating students are mostly persons of the middle age, which are much less oriented on cheating.

Therefore the process of study, practises and examination should be built basing on completely other principles than in „normal" university courses oriented on young students. Ideally it should be dependent on the person accessing the subject in case the same course can be taken by both re-educators and by university students (in case it is, for example, included into a graduation program).

C. Reasons

Obviously, the state-wide educational system can be built basing on completely different approaches, for example by:

1. involving older people into the course in universities using "open university" system;
2. building courses for unemployed;
3. organising centres of re-education where subjects are given like in university, but purely for older people considering their ability to learn;

… and many others.

The paper would like to propose an idea of organising a virtual educational environment accessible for any citizen of the state. This idea has a set of advantages discussed below.

First of all the ideal educational process should gradually upgrade skills and knowledge of employers and be proactive rather than reactive. It should not be oriented just on unemployed persons, as this status should be seen as extreme and avoided in all possible ways – including providing a possibility to learn for those who still work, but already fill some lack of knowledge. The citizen should have a possibility to obtain new knowledge (new profession) before s/he will be thrown away by the current employer, not exclusively after this will happen.

Obviously different persons has very different abilities to study and this ability doesn't necessary evolve with the age. Vice versa, for many people their ability to learn will decrease the older they become. Therefore it is very difficult to propose to put such different people into the same class. Consider also extremely different backgrounds, systems of values etc for young and older people and you will see that combining different age people in the same class is not a good idea especially in the massive study process, which is practised in colleges and universities.

Reconsider also the fact that different people have different abilities to learn, so slow learners will slow down the education process in respect of quick learners. All this demonstrates that we need an environment, which can be adaptable to the concrete persons needs. Those needs include ability to learn generally and a level of understanding of one or another topic (i.e. if one person basing on his/her background could find a certain topic to be obvious and easy to learn, then another could struggle during quite sufficient time before it will become clear) and many other person dependent parameters. An ideal system for such different students will be the virtual education environment capable to obtain context information and serve knowledge quick enough to move forward and slow enough to keep a person understand the topic.

Secondly the country wide system should be oriented on employers to constantly upgrade their knowledge, which means that they rarely have time to visit lectures during 8 hours per day. This also includes a possible inability to attend lectures during a certain time that suits the lecturer and so makes extremely hard to organise in the face-to-face communication (i.e. read in the off-line mode). The only exception again is the virtual environment when the course (or any part of it) can be taken when the person has enough time after work, social or family duties. It is quite possible that s/he can spend just 15 minutes per day or several hours per week so all other ways to organise the education process will fail and e-education will be the only still able to support it.

Finally we should consider the main goal of the process. The system should be state wide and accessible for all citizens, which means that there could be millions of attendees. It is hardly possible to collect of them in any special places even after clustering them into towns and groups and is rather easy to be organized basing on virtual environment. It doesn't mean that it is very easy as we need to ensure telecommunication channels etc, but will be much more feasible than other methods.

Consider also the fact that it is much easier to build a course once for millions rather than replicate it in each schools risking to loose in quality. Notice that individual courses are greatly dependent on the teacher knowledge and skills and so will not be equal among different areas or groups of people. Here we arrive to another strength side of the e-education environment. It is hardly possible to ask an outstanding professional to host a lecture for a very broad auditorium, to host a set of those – due a restricted ability to travel and attend meeting of both speaker (primary) and students (secondary). Here the primary element is the main restriction, but the secondary one is much more massive (predictable 30-40% of attendees will not be able to visit the meeting especially if those happens periodically) and so much more important to the government, which tries in our case increase the level of knowledge of millions citizens and affect the economy in such a way.

D. Components

The first interface that gets most attention is the consumer interface – in our case interface for the citizen / student designed for learning process. Obviously it should contain a log-in service provided by a security server and the integrated standard environment via which a course is selected. The integrated environment is a portal to start, continue and complete different courses, find marks etc. That portal is based on several underlying layers like course-warehouse, adaptation system, context layer and personal scorecards and performance meters.

As our e-educational system is not designed just for students, but has much broader, state wide meaning, it should also contain the organizations interface. This should provide a possibility for organizations to affect the system in order to adopt it to the current business needs using, for example, the following components:

1. Statistical information on taken course, acquired marks and so forth – this part is designed to show how much potential work-force is available for the business in a certain technology area and could be consumed making a decision to migrate /start to employ) one or another technological platform, building business in certain area and so forth;
2. A possibility to provide advertising information targeted to people taken certain courses. This will allow to student see what courses could potentially provide them a new work-place and for companies to find more employees.
3. A person-business communication portal via which a certain person could give access to his/her result by courses to a company. It can be used in order an organisation would like to verify first of all knowledge or different candidates for an open position either as a preliminary check or in absence of professional in a certain area able to make correct candidates selection.
4. A course creator-business communication portal, via which a company could order creating a certain courses to grow up potential workers or demand to include into existing course some parts crucial for the current state of the art in this area (technology, business etc)

Finally the system requires a course creators interface exposing standards and templates and including forms allowing to upload courses and provide other necessary information that the e-education environment could allow to define in order to adopt the general student interface to this particular, specific course. Notice that so far we aim to describe the system where exams are either never done or automated, so the system will not require a constant involvement of professionals to provide online lectures via web-cams or revise exam answers in order to decrease the overall running cost of it. At the same time it doesn't mean that such element should not be ever added – it just should be carefully considered and planned before including it into the standard package offered to students.

More details on the model and interconnection between those components are provided in the next chapter called "Model".

E. Responsibility

It is a common mistake to assume that the only responsibility the organization creating such portal has is a responsibility to make it work and to ensure it correct and constant service. This task is very important but not the most crucial as a temporary system unavailability will not be as harmful as a lost of private data that the system collects during its work.

The most obvious type of private information is the set of marks obtained during taking courses if the corresponding functionality exists in the educational system. The next one is the log-in system data, which the system consumes and so could incorrectly expose or illegally collect. Thereafter the list of courses that a person has taken should be also included into the sensitive information list. Finally modern educational systems contains a lot of artificial intelligence like context-aware logic that does collect a lot of information about the person and produce even more analysing both his/her behaviour and other and so arriving to some conclusions – all such patters, clusters and knowledge should be considered as extremely sensitive and therefore highly confidential information

Summarizing all previously said system owners / holders have a clear responsibility to protect privacy data and carefully use within the system to ensure that those will neither be stolen nor made available to any other third party.

F. Conditions disabling implementing the proposed kind of system

Finally we will recollect cases when the proposed ideas of massive and constant re-education of country citizens based on virtual education environment cannot be applied

1. First of all it cannot be applied for a variety of subjects requiring hands-on skills to be obtained in special centers equipped by correspondent mechanisms or having certain conditions. Consider for example heavy industry, builders requiring an area to practice building walls, painting etc.

2. Secondary the system requires a sophisticated development of telecommunication channels and so certain involvement of ordinal people into electronic data interchange, having web and computer skills. Similarly government should have enough professionals to be involved into organizing the process, maintaining hardware and software, preparing courses etc.

3. Government should reach a certain sophistication level to understand the advantage of the state wide education process, which is not direct (money paid for the course) and will arrive via increase in economy and taxes. The weak or poor country rarely can afford executing such sophisticated programs.

III. MODEL

A. Hosting environment and security

Consider one particular country - Estonia as an example in order to set a starting condition the state educational system can evolve from. Estonia has a central informational system developed by the government via which different services are exposed to citizens. Mostly those services allow receiving information on the citizen from different state institutions like tax office, medical insurance register and state hospitals, national motor vehicles' register and so forth. Recently the list of services started to expand and include now possibility to post documents that are electronically signed and so are mandatory for all state departments to accept and process. The information system has a single entry point, but the authentication methods vary. Several years ago the typical method to identify was using a bank log-in system, that uniquely identify the person and so can be used to obtain the person identification number as well and consequently log into the state system. Lately the system started to change as the government decided to promote citizens' identification cards, that contains passwords (pin codes) and so are mostly used nowadays for both the state information system and for banks as is the best way to identify yourself – if another person passwords using sometimes can be seen as an act to be punished in administrative way then using stolen id-cards is a crime by the law as those are equal to passports.

This system can serve as a starting point as allow identifying a citizen (student) and provide a set of services, one of which can be the educational one.

B. Student's interface

The most noticeable part of the system is the end-user (student – citizen) interface

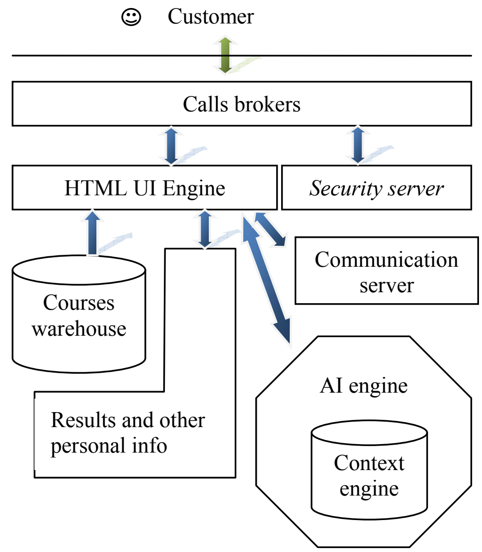


Figure 1. Students interface

Consider the figure 1. Here the customer (student / citizen in our case) communicates to the system via the wire like Internet using a desktop or mobile. The system gets the request and the calls broker identifies it and checks whether the session was already open or the persons should be identified. In the last case the call is rerouted to the security server, which is denoted on the figure by the italic in order to highlight the fact that the identification system could actually locate outside the e-environment scope, so it just consume some other system's identification process. The security server can use also biometrical data [8, 9] for participants' identification [10] or even a constant check of the data flow based on biometrical patterns [10]. After the log in process the call is routed to the main HTML UI Engine of the system that constructs the student portal's user interface basing on his/her privileges, earlier interactions and available courses. Here the html interface is used as a standard way to provide web pages, although it can be implemented using whatever software development language including Java, Flash, Active server pages and so forth. The UI engine relies on several subparts:

1. Courses warehouse: a store of all courses added into the system including all kind special elements of each course like presentations files, structure logic, animation or video files and other parameters defining the feel and look interface within predefined standards

– parameters making the course typical (standards) and different (course custom settings) from others.

2. Results and other personal information: this part contains current and historical results and progress per taken courses including acquired certificates and communication to other parties (students, organisation etc)
3. Communication server: this kind of server provides communication services like forums and chats including communication between students or communication to organisations, so advertisement. As earlier was described via this service the student could publish own certification and courses information to chosen companies and vice versa.
4. Artificial intelligence engine (AI engine): this part of the system add intelligence into the environment able to propose course to be taken, vary speed of explanations or modifying the course by skipping beginners parts if needed, adding extra references on other courses and so forth. Usually this system is empowered by the context engine allowing producing a better description of each student rather than its identification number does and so enables more flexible intelligence than usually. The context-aware systems are evolving at the moment very quickly and are very popular among information systems creators.

C. *Busniess community interface – external data consumers*

Here a business community interface is presented as they will be very much interested in results of the education process as the paper described earlier.

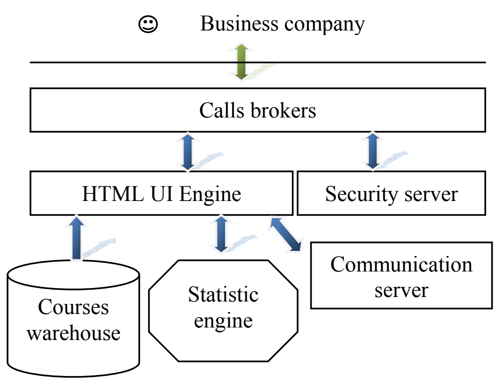


Figure 2. Business community interface

Here the interface is built using the same core as for the previous one, but user interface engine for this group of users generates completely another interface as their privileges, interests and goals are sufficiently different from the students group. The business community portal presents general information on available courses, statistical information and has a sub-interface to communicate with other groups of users.

First of all, the portal offers a service to collect statistical information of e-education students by adding into the system a dedicated statistic engine. This analytical engine is able to traverse all marks obtained by students, courses they started or completed and could provide quite interesting information into addition to the current market one, which can indicate the future trends and allows planning business processes focused on the work-force with a better predictability rate.

The interface contains also a communication services allowing posting advertisements targeted to students and requests to courses creators and system holders. The communication process is bi-directed and creates the collaborative environment allows improving both courses and students understanding of what others (that the current employer) business organisations could demand in order to secure his/her future by expanding knowledge in the most demanding areas.

D. *Courses' authors interface*

The last interface to be described in this paper is courses creators one. This interface is very similar to earlier described although is not identical.

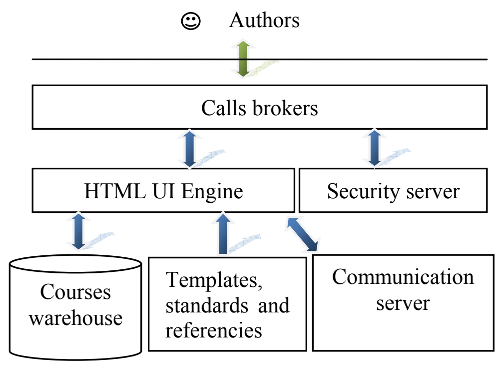


Figure 3. Courses' authors interface

First of all authors the only group of persons able to define and upload courses and other materials into the correspondent warehouse. Besides the do consume special information defining how those courses should be created basing on templates, standards and other guidelines published by the e-education system owner. Those materials are designed to standardise courses and make look and fill of those similar in order to increase students' ability to consume those: owns taught interface or way to learn can be used for any course of the system.

IV. CONCLUSION

The e-educational system described in the paper is the crucial components of ensuring the stability and financial health of any state. The workers productivity was always one of the major factors defining the success of the state industry and innovation and therefore requires a close attention and funding. The new era is even more challenging for both citizens and government as rapid changes makes them ensure in the future and the proposed e-education system is a simple, efficient and not expensive way to upgrade skills. At the same time it does require certain sophistication of the state, citizens web knowledge and telecommunication channels and therefore can be recommended to be executed first of all either in highly developed countries or states' unions, like for example European Union, that both acquired the required level and have state institutions able to fund this program.

REFERENCES

[1] D. Mowery, R. Nelson, B. Sampat, and A. Ziedonis, "Ivory Tower and Industrial Innovation," *Stanford Business Books*, California: Stanford, 2004.

[2] K. Fischer, "The university as economic savior: as higher education replaces industry in some cities, colleges face both support and unrealistic expectations," *Chronicle of Higher Education*, 2007.

[3] J. Youtie, and P. Shapira, "Building an innovation hub: A case study of the transformation of university roles in regional technological and economic development," *Research policy*, vol. 37 (8), pp. 1188-1204, 2008.

[4] D. S. Siegel, D. A. Waldman, L. E. Atwater, and A. N. Link, "Toward a model of the effective transfer of scientific knowledge from academicians to practitioners: qualitative evidence from the commercialization of university technologies," *J. of Engineering and Technology Management*, vol. 21(1-2), pp. 115-142, 2004.

[5] B. Readings, *Education / Teaching*, Harvard University Press, 1997.

[6] E. González-Agulla, E. Argones-Rúa, C. García-Mateo, and Ó W. M. Flórez, "Development and Implementation of a Biometric Verification System for E-learning Platforms," *EDUTECH, Computer-Aided Design Meets Computer-Aided Learning, IFIP 18th World Computer Congress*, 2004, pp. 155-164.

[7] O. Guven, S. Akyokus, M. Uysal, and A. Guven, "Enhanced password authentication through keystroke typing characteristics," *Proceedings of the 25th conference on Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications*, 2007, pp. 317-322.

[8] A. K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 14 (1), 2004, pp 4-19.

[9] A. K. Jain, L. Hong, and S. Pankanti, "Biometric Identification," *Communications of ACM*, vol. 43 (2), 2000, pp. 91-98.

[10] D. Kumlander, "Soft Biometrical Students Identification Method for e-Learning," *in Advances in Computer and Information Sciences and Engineering*, Springer, Netherlands, 2008, p. 114 – 118.

A Context-Aware Biometric Security System for e-Learning Applications

Deniss Kumlander
Department of Informatics, Tallinn University of Technology,
Raja St.15, 12617 Tallinn,
Estonia
kumlander@gmail.com

Abstract-**The paper describes a context-aware security system for identifying users mostly during the exam phase in the e-Learning environment. The system employs soft biometric characteristics of the individual and environments parameters in order to capture patterns and later applies those on reactive base identifying the person participating in the exam including external help that could be offered to him illegally from the teacher point of view. The paper discusses how the system can be built and what methods should be implemented within the system prior to the examination in order to increase reliability of the system.**

I. INTRODUCTION

Importance of an electronic learning as an alternate solution to the traditional education has sufficiently increased recently with development of electronic channels and world globalization [1, 2]. There are millions of students attending e-courses in developed countries and this number is constantly growing. Financial crisis, evolution of technology also demand a constant update of skills for nearly every member of modern society on the constant base, which also adds a lot of additional students and moves the e-learning process out of young-people oriented scope.

The main problem that arises in the e-learning, especially in the examination process is students' identification [1]. There are several methods that can be applied to solve this problem including certified centers and "hard" biometrical methods as requirements to attend the examination using web-cams. At the same time those approaches mostly do require using some special hardware, which is sometimes either unavailable to students or they are not technically skilled enough to install and use those, so in result they tend to afraid that. Summarizing all previously said, the current students' identification methods sometimes prevent people from using e-courses [3]. Therefore it is important to develop such security identification subsystem, which will be transparent for the end-users of e-courses, but will be good enough to identify users and warn teachers in case of any suspicious activity. The system should be flexible enough to adapt to very different environments users can be consuming the e-portal from.

II. E-LEARNING ENVIRONMENT – THE PROBLEM STATEMENT

The problem statement for identifying students and using the online e-learning environment greatly depends on the auditorium to which e-course is targeted. In other words, before applying one or another security system you should consider "Who and why": who is studying and why - what are their goals during the study process. The main problem we are going to research in this paper is a security risk that the person doing the exam is not the one who gets the final mark.

III. BIOMETRICAL MARKERS

Biometric is a relatively novel approach to the computer security that uses either physical or behavioral person characteristics to identify that person. The major idea is to learn computers to identify persons as people do it in the everyday life. The most well-known system of identification is finger prints. In fact, it was used much longer than electronic computers do exist at all. Nowadays, there are much more types of different security approaches for computers relying on biometrical methods. Consider for example face recognition, eyes' patterns, signature or voice recognition [4, 5]. Unfortunately most biometric approaches require special hardware systems and therefore are not good enough to be used in e-learning as rarely students have such devices or are ready to buy those.

The biometrical security system is relatively easy to implement and use. It is possible to vary the factors to be measured and users will be verified by. The system can start from some basic measures and could evolve by implementing more and more features. The general approach to storing, capturing and using data is very similar to other computer based recognition systems. Typically there are three basic elements as it is demonstrated on Figure 1. The first part is training block responsible for registering users in the system by producing patterns for each user during users' communication with the system. The sensor produces raw data and an analyzer generates patterns, which are:

1. Bounded to a particular measure, like time intervals;
2. Pre-processed and stored in a compact way ready for later use.

The registered patterns are saved into a patterns database, which is the second part of the system. After that, when user logs into the system and uses it, the third element of the system activates. During the matching step, a user pattern is once again captured via the same sensors and compared to the saved

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_84, © Springer Science+Business Media B.V. 2010

one. If those are different then the system issues an alert to the environment administrator on this user. The similarity level of the compared patterns is normally tuned by setting a critical level, lower which patterns are called different. The system should adjust patterns constantly in case individual's behavior is changing as s/he become more familiar with computers etc.

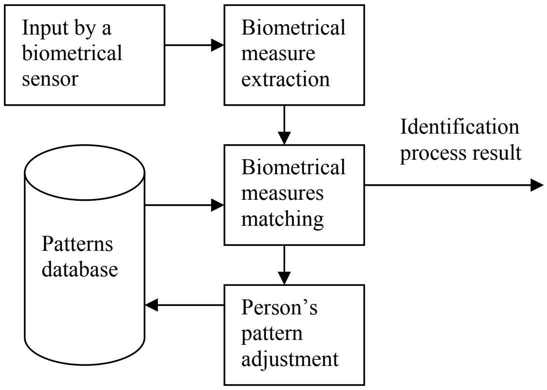


Figure 1. Adjustable biometrical identification systems architecture [3]

One of the simplest and so the best method to be used in such biometric security system is keystroke patterns, where the system monitors first of all the keystroke typing dynamic. It is well-known among telegraph operators that writing/typing dynamics are assumed to be unique to a large degree among different people [6]. Actually the keystroke typing dynamic can be measured by the following sub/characteristics [2, 6]:

1. Duration of the keystroke, hold time;
2. Latency between consecutive keystrokes;
3. Overall typing speed;
4. Habits by using additional keys on the keyboard like typing numbers via number pad part;
5. Frequency errors and backspace vs. delete key using; and some others

The keystroke recognition systems research is started in 80s [7] and includes a set of specific recognition methods like [2]:

1. Statistical methods [7, 8, 9];
2. Artificial intelligence methods / machine learning and data mining, including but not restricted by neural networks, graphs and Euclidean distance metrics, decision trees etc [10, 11];
3. Generic algorithms [12] and fuzzy classification methods [13].

IV. Context-Aware Systems

A context aware software development principal defines the computer reaction can be organized using more than just direct data input. It could and probably should consider circumstances under which the input is defined, sense the environment in which the request is formulated and so forth. In the result the system response will be different for the same request depending on those extra (request context) parameters.

Some authors [14, 15] say that this behavior can be described by the system ability to adapt to specific contexts under which the system operates or reacts on the request and ability to collect and recognize the context information.

Consider for example web applications, which do normally react on the request, but could collect and modify behavior depending on the following extra parameters:

1. Previous requests that are not considered as direct input to the current one. For example so far collected preferences of a user without asking him to define those directly like previously made searches in case of search engines etc.
2. Internet browser type. A request made using IE browser should be responded differently in compare to Firefox as those browsers standards (JavaScript, HTML) etc are different so the response will be rendered differently;
3. Accessing device – obviously the page should look different in case the user accesses the page using mobile version of the device due space, keyboard etc restrictions;
4. Vary the response depending on geo location of the requester (country, time-zone, language);
5. Connection speed – in order to provide either a short and restricted answer or expanded and large one;

… and so forth.

The context-aware approach is used to increase relevancy of the response (service) provided by the system and target to bridge the gap between formal and [16] human-human type interaction. Generally context-aware application contain 3 types of sub-activity elements: they should detect context, interpret it and react on it, although the first activity is sometimes delegated to other systems (alternatively it can be said that the system relies on other one to collect context and provide it and so depends on the front-system in terms of how well it can adapt as the smaller context is the less flexibility the system could provide).

V. A Context-Aware Security System for e-Learning Based on Biometrics

Here, a system combining both biometric measures used for identification of users and context-aware principals will be described. This part of the work is the next step in our research of applying "soft" biometric characteristics in the identification process [3], where "soft" means such characteristics recognition, which doesn't demand any special hardware to be used – we use only a standard set. For example the keystroke one can be obtained using PC with a keyboard and embedding client-side software into the portal pages, via which the user interacts with the system. This chapter proposes how soft biometric measures centric system can be converted into a security (identification) subsystem by using context-aware software development principals. Besides the security system is sensitive to the amount of information on biometrical characteristics flowing through the system and is the more precise the more data is collected. Therefore some extra steps should be implemented applying the system prior to the exam phase to increase the reliability of this sub-system.

First of all patterns should be recorded and verified prior to using, i.e. identifying students. Depending on the environment, purposes and teacher/student relations the process of collecting biomedical information can be either hidden or visible to students. Normally data would be recorded during the study period and the method (markers etc) greatly depends on how the exam will be organized and so the plan should be compiled before implementing such security subsystem [3].

Thereafter we do propose that at least several biometrical measures are used and at least the keystroke dynamic, used keys and used words should be implemented to make the system acceptably reliable. It is well-known nowadays that person's vocabulary greatly depends in his background, skills, intelligence and family s/he comes from. Therefore it is possible to distinguish persons basing on typical words and phrases s/he is using and nowadays it is a method that was quite successfully applied to analyses different authors' texts (novels) to identify authors of some unsigned texts.

Next, we recollect different exam scenarios to see how the proposed system will work and what elements are designed to prevent incorrect (illegal from teacher point of view) way of passing exams.

It is possible to distinguish two types of exams from the students' point of view. The first one is "doing exams in the class room equipped with computers" and the second one is "participating in exams from home". In both cases e-learning portal pages responsible for examination process should contain certain software elements to constantly monitor the user input via answer s/he is giving to the system on exam questions. There is a huge difference between bio-measures generated by a person writing an answer by him- or herself versus any version of using somebody's help. The last case can vary, but still can be easily recognized in most cases.

First of all an answer can be obtained from somewhere else (e-book, others posts or even from a chat having somebody online helping you). The most often pattern here will be a copy-paste scenario where the answer is not typed in, but is copied into the answer box instead. Obviously it can be easily recognized and an alert issued.

The second case, when the person is using a help from another person physically presented in the same room, via the mobile etc. Here two potential patterns mismatching could occur. First of all a words using pattern will be broken, as the person is not using his/her own expressions, words and so the distribution of words will be considerable different. Secondly the rhythm will also be different. Although the second person also do have pauses to think, still s/he will dictate faster than the person will write having already formulated the answer fully – earlier than the writing one will be able to fill the answer text box. Therefore the rhythm will sufficiently differ from the one that writes own text including the use of backspaces (delete button).

If the exam demands code writing or formulating an answer in any special symbolic language then the pattern detecting part should consider this as an extra context on the flow. Obviously the pattern will sufficiently differ here from a free text writing (like writing an essay) and is greatly depend on the nature of this symbolic language: on both syntax and semantic. Although it sufficiently narrows the number of used lexical constructions, it also sufficiently improves recognition of beginners as the skill of high speed writing correct constructions, number of post writing corrections (forced by the syntax parser) etc will indicate the individual's level of knowledge. Unfortunately it will also slightly vary depending on personal attitude to writing the code: some prefer to write quickly producing a raw code and thereafter will fix errors. Others would like to think everything through and only then start writing the code producing more or less clean result. Therefore there is a need to implement some extra activities to both increase reliability of patterns and use such differences in the attitude. Some small exercises should be designed and offered to students during the learning period. It will allow both measuring the attitude of each one and measuring the skill evolution trend. Obviously if either the attitude will change or student progress level differs dramatically from what s/he had during the last exercise then an alert is raised.

The well-known problem of applying biometric measures is a sufficient variation of those in different environments, like home, train etc, from which s/he connects to the e-learning portal. In systems proposed so far those environments were equivalent, in other words this factor was completely ignored. We do propose applying the context-aware system principles and will try to recognize different environments and so write different patterns for each. It can be done by reacting on different external systems parameters posted together with the request or by analyzing patterns: the current and stored in the database. If patterns differ sufficiently during the capturing then we deal with different environments. In order to avoid cases when different persons recognized by the system as one with different patterns we need to monitor frequency of using one or another pattern, as occasionally used patterns can be ignored matching during the exam and so issuing alerts. Anyway the patters based analysis should be secondary and external system parameters should be primary. It is possible to recognize:

- Device type;
- Browser type
- Drivers (and so identify the keyboard)
- Connection speed

Clustering by those parameters will allow identifying most environments like a PC at home versus a PDA on the way to university (work) in a train: obviously those are different devices, different connection speeds and different browser types.

It is important to see additional characteristics beyond the technical environment identifying the context. The history of system-user communication is also a part of extra information that can be used to increase intelligence of the system and so adaptation. This method requires clustering students and keeping a database of answers, progress etc. for each student. For example, the approach asks to use intermediate knowledge checks during which his/her correct answers are saved. After that some of those questions are presented during the main exam and if the student doesn't know the correct answer any longer then an alert should be issued. It is possible to check the vocabulary of the answer as well and see how the student has formulated answers of the well-known (for him/her so far) questions [3].

VI. Responsibility

The collected patterns should never be exposed to the external systems. It is a part of private and so sensible information and the company responsible for the system has a full responsibility of keeping this highly secret.

In order to fulfill this requirement the central patterns security sub-system should expose a functionality that can be embedded into the html page and so will download client side scripting for capturing patterns. Notice that the system could also collect extended context information about the current session like the browser agent used (in order to distinguish between „full" browser and the mobile one, IP address assuming than difference in IP addresses could mean another keyboard, another (more or less noisy) environment and so forth) here instead of relying on any other subsystem.

In addition, this central approach will also guarantee that the pattern stored within the system will be constantly updated accumulating the old one and the new, recently captured.

VII. Concluding Remarks

It will be incorrect to rely only on the proposed approach building a security system for identification of students during an examination phase of e-learning. The soft biometric measures and earlier mentioned approaches to increase their reliability do not guarantee a full identification of students by themselves, but those do provide enough information to apply extra methods for students' identification in case a biometrical system alert is issued. For example the exam can be divided into two parts. The first one is an online, long and complex exam. The second one is a very short face-to-face check to compare the long exam answers with the person knowledge to verify the exam end results. The face-to-face exam can either happen only if an alert was issued or should be done using more questions versus very few questions one if no alerts were produced for this particular student.

The security subsystem described so far should not be implemented mechanically. For example it does not include a protection subsystem against students that are using course materials when they are not supposed to do that. At the same time it is always possible to organize questions so, that they will not require direct citations, but instead will require knowledge acquired during the course in order to formulate correct answers.

VIII. Conclusion

A student identification process is a key question for modeling, implementing and executing any e-learning environment. An incorrectly implemented security module can either fail identifying persons allowing obtaining certificates for wrong individuals or sufficiently decrease the e-learning system accessibility and audience. The paper proposes an identification sub-system employing individual characteristics of a person (biometric) and a context-aware software development methodology. The system records patterns using several biometric measures. Those patterns are used later during the examination phase recognizing the difference between past and current pattern and issuing a security alert if the different is over predefined acceptance level. A context is an essential part of the system. Considering the context the system can capture several patterns in order to distinguish between different environments a user acts in and history so skills' evolution trends can also be applied verifying the person. The paper does propose in addition several extra activities, which are required to implement a biometric context-aware system correctly as the identification process greatly depends on quality of patterns captured before exams, i.e. during the entire study process.

References

[1] E. González-Agulla, E. Argones-Rúa, C. García-Mateo, and Ó W. M. Flórez, "Development and Implementation of a Biometric Verification System for E-learning Platforms," *EDUTECH, Computer-Aided Design Meets Computer-Aided Learning*, IFIP 18th World Computer Congress, 2004, pp. 155-164.

[2] O. Guven, S. Akyokus, M. Uysal, and A. Guven, "Enhanced password authentication through keystroke typing characteristics," *Proceedings of the 25th conference on Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications*, 2007, pp. 317-322.

[3] D. Kumlander, "Soft Biometrical Students Identification Method for e-Learning," *in Advances in Computer and Information Sciences and Engineering*, Springer, Netherlands, 2008, p. 114 - 118

[4] A. K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition,", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 14, no. 1, 2004, pp 4-19.

[5] A. K. Jain, L. Hong, and S. Pankanti, "Biometric Identification", *Communications of ACM*, vol. 43, no.2, 2000, pp. 91-98

[6] J. Illonen, "Keystroke dynamics," *in Advanced Topics in Information Processing Lectures*, 2003.

[7] R. Gaines and W. Lisowski, S. Press, and N. Shapiro, "Authentication by Keystroke Timing: some preliminary results," *Rand Report R-256-NSF, Rand Corporation*, 1980.

[8] R. Joyce and G. K. Gupta, "Identity Authentication Based on Keystroke Latencies", Commun. ACM, vol. 33, no. 2, 1990, pp. 168-176.

[9] I.Sogukpinar and L. Yalçin, “User identification via keystroke dynamics”, *Ist. Univ. Journal of Electrical and Electronic Engineering*, vol. 4, no. 1, 2004, pp. 995-1005.

[10] S. Bleha, C. Slivinsky, and B. Hussien, ”Computer-access security systems using keystroke dynamics”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, 1990, pp. 1217-1222.

[11] Y. Sheng, V. V. Phoha, and S. M. Rovnyak, “A parallel decision tree-based method for user authentication based on keystroke patterns”, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 35, no. 4, 2005, pp. 826-833.

[12] E. Yu and S. Cho, “GA-SVM wrapper approach for feature subset selection in keystroke dynamics identity verification”, *Proceedings of the International Joint Conference on Neural Networks*, vol. 3, 2003, pp. 20-2253 – 2257.

[13] B. Hussien, R. McLaren, and S. Bleha, “An application of fuzzy algorithms in a computer access security system”, *Pattern Recognition Letters*, vol. 9, no. 1, pp. 39-43.

[14] P. Chang and A. Gul, "Towards Context-Aware Web Applications," *7th IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS)*, Springer, LNCS 4531, 2007, pp. 239–252.

[15] W.Y. Lum and F.C.M. Lau, “A context-aware decision engine for content adaptation,” *IEEE Pervasive Computing*, vol. 1, no.3, 2002, pp. 41–49.

[16] J. Pascoe, “Adding Generic Contextual Capabilities to Wearable Computers,” *Proc. 2nd Int’l Symp. Wearable Computers*, IEEE CS Press, Los Alamitos, Calif., 1998, pp. 92–99.

Creating an Immersive Virtual World through Integration of 2D and 3D Technologies to Implement E-Learning Curricula for Middle School Students

Coton, C., Shegog, R., Markham, C., Thiel, M., Peskin, M., Tortolero, S.
Center for Health promotion and Prevention Research
University of Texas School of Public Health
7000 Fannin Street
Houston, TX 77030 USA

Abstract- **E-learning applications are becoming ubiquitous in the classroom. However, they are more commonly implemented within the context of flat, 2D content presentation and testing. In some cases an immersive, 3D learning environment may be more compelling, yet implementation of these systems is often costly and resource-intensive, requiring high-cost development tools and hard to find skill sets. This paper describes a low-cost innovative development and implementation of a successful LAN-based HIV, STD, pregnancy prevention middle school curriculum that featured a 3D gaming interface, 2D tailored activities, and video components. The resulting 3D virtual world containing embedded, pre-existing, 2D activities and video at activity stations met criteria for functionality and usability while meeting educational objectives and cost constraints.**

I. Introduction

While E-learning applications are ubiquitous in the classroom, they are more commonly implemented within the context of flat, 2D content presentation and testing. In some cases an immersive, 3D learning environment may be more compelling, yet implementation of these systems is often costly and resource-intensive.

A successful comprehensive HIV, STD, pregnancy prevention curriculum for middle school students [1, 2] has been adapted as an abstinence-only curriculum [3]. These curricula provide a unique approach to the application of technology in the classroom by integrating group-based and computer-based lessons.

The purpose of this development effort was to meet the educational objectives of broader CDC-funded HIV, STD, pregnancy prevention studies, while addressing the requirements for creating a cost-effective development process. An approach consisting of embedded, pre-existing, 2D Adobe Flash™-based [4] activities and video at activity stations within a 3D virtual world met both educational objectives and cost constraints. Low-cost development tools, including the game engine, 3D modeling and raster graphics tools were used to develop and deploy the e-learning system to the classroom.

II. Methods

The approach taken to establish virtual world development workflow was subject to selection of a 3D game engine and 3D modeling software on specification criteria of function, extensibility, and cost. Common commercial components were used wherever possible to implement capabilities such as network connectivity and data storage, in order to keep development costs within constraints.

A. 3D Game Engine

Game development software selection was driven in large part by the need for flexibility and extensibility, as study requirements could radically change on a week-by-week basis. A 3D game engine with plug-in capable architecture and a C-type DLL interface were vital to allow the quick development or purchase of components providing functionality to meet study requirements.

Minimum engine graphics requirements for target computers were also of concern. Whereas study requirements provided the advantage of fixed hardware specifications, computer selection was still constrained by cost, and target machines' graphics processor memory was limited to 128 MB. Strong consideration was given to game development systems with integrated world and script editors, in order to provide an efficient workflow during system development. Whereas fully integrated development environments are rarely found in low-cost game development packages, a system with an integrated world editor was considered a minimum requirement. Conitec 3D Game Studio Pro™ [5] met study functionality and cost requirements.

B. 3D modeling

Requirements specifying low-complexity models for the virtual world allowed for the selection of lower cost 3D modeling applications and utilities. Complexity of models in the virtual world was low, with an average of 1500 to 2000 polygons per model, generally limited to buildings and props representing an outdoor shopping concourse with terrain and sky. The architectural 3D modeling tool SketchUp™ [6] was used to create buildings and many of the props within the virtual world. Model files created with SketchUp were

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_85, © Springer Science+Business Media B.V. 2010

exported to the commonly used 3D Studio Max™ format, which allowed the use of low-cost utilities to perform any needed operations on the models' polygonal mesh. Other low-polygon props such as signs and 3D text were more easily created with the shareware modeling tool MilkShape 3D™ [7]. As the targeted mid-level graphics processor required a minimalist approach to 3D model polygonal detail, every effort was made to create models with the lowest possible polygon count while maintaining visual integrity. Software such as Deep Exploration Standard™ [8] was used to reduce the polygonal count and to manipulate polygon normals of each model. As SketchUp was designed primarily to create architectural models that support 'walk through' capabilities, model meshes created with the software contain two faces per polygon. To further reduce runtime graphics processor load, it was often necessary to remove one face from the mesh of models not facing students during play. Only simple materials were used for each mesh; the use of shaders was avoided. When the architecture of each model was complete, the shareware utility Ultimate Unwrap 3D™ [9] was employed to map UV coordinates to its polygonal mesh in preparation for assigning textures or "skinning" the model. Textures for each model were created as bitmaps within Adobe Photoshop™ [10]. As many building and prop models required transparent components to simulate glass in windows and doors, the 32-bit Targa™ format was used for model texture bitmaps. The Targa format provides 24 bits per pixel to store image data, with an 8-bit alpha channel reserved to create the appearance of partial transparency within an image.

C. Virtual world creation

Using the world editor in 3D Game Studio, a terrain mesh was first created, skinned with a bitmap and positioned at the origin in the workspace. Locations of activity stations to be represented by buildings or outdoor props were laid out with cubic primitives provided within the editor. Hollow cubic geometry representing sky was then created using a 6-sided graphic of a nighttime sky and positioned over the terrain, again using world editor utilities. Using the world editor, light sources were then strategically positioned above the terrain and near proposed building locations to simulate an outdoor nighttime experience. Ambient sound sources were similarly positioned. In order to address limitations on the target graphics processor, lighting within the virtual world was calculated at compile time and remained static. Unlike many virtual worlds, player movement was constrained to a relatively limited area in order to maintain students' focus on task while maintaining the context of playing a game. To achieve this, invisible architecture was placed within the virtual world to set boundaries to players' movement. To eliminate the complexity involved with creating and animating player avatars, a first-person camera view was used for each student. Once the starting position of the default camera view was established, placeholders for architecture within the world were replaced with 3D models.

D. Database connectivity

Data storage and retrieval capabilities were required for runtime configuration of the virtual world, and for tracking student interaction at runtime. The study required frequent educational content changes due to changing student grade-level targets and different intervention approaches. Often, content changes were required in the field, post deployment. To address these requirements, the system used data stored in a Microsoft Access™ database to determine the curriculum content to play at each activity station, as well as to store students' quiz responses and time on task for each activity. A freeware data access library written by members of the 3D Game Studio community was used to implement access to the database. Source code written in the game engine's scripting language called data access related functions via the library's DLL interface, employing Structured Query Language statements to read and write data.

E. Networking

In a typical classroom, students would concurrently participate in single-user instances of the virtual world. Some activities would require students to post responses to short-answer activity questions, which could be viewed by the study facilitator, using a simple reader application installed on the facilitator's computer. After students completed a virtual world session, study facilitators would engage students in a discussion focusing on their responses entered from within the virtual world. In order to implement this functionality without the development of additional software to implement instant-message client functionality, the system employed remote database connections over a Windows network. The virtual world was installed in a Windows XP™ environment on each student's computer, while the study facilitator's computer needed Windows Server 2000™ to handle the connection load. Each XP client connected at startup to a shared directory on the facilitator's server, where students' responses were saved to an Access database during virtual world sessions. As students entered responses to discussion-related activities, the responses were copied anonymously to the remote database, and were displayed to the facilitator via the simple reader application.

F. 2D content integration

The study e-learning activity content was developed by a 3rd party commercial firm in 2D Flash format. Windows Video content was also used in the curriculum. The study curriculum required this content to be made available at activity stations during sessions within the virtual world. Whereas the 3D Game Studio game engine provides methods to render video on textures within a virtual environment, results using these methods were unsatisfactory in terms of video quality. The system was therefore designed to execute interactive Flash activities and video segments as separate processes. All Flash-based activities were first converted to self-playing, executable projectors. During a virtual world session, whenever students came within the proximity of the location of an enabled activity, collision detection scripts would retrieve data referencing the appropriate file containing an activity or video, and would spawn a separate process to play the activity.

This translated to the following lesson functionality. During a session, as a student moved close to a glowing prop representing an activity station, the 2D Flash activity or video would begin. Each activity's window was sized to cover that of the virtual world, movement within the world was temporarily suspended, and mouse control would pass to the activity window for the duration of the activity. A segue at the conclusion of each 2D activity presented verbal direction to the next activity. After all activities were completed, students were allowed to freely explore the world, including an "Arcade" where previously completed activities could be replayed.

G. *Beta testing*

A single group usability study was conducted in facilities at the University of Texas, Houston Health Science Center with a convenience sample (n=19) of Houston middle school students representative of the school districts demographics: primarily minority (52% Hispanic) and 68% female. This sample tested the 8^{th} grade lessons. Participation was voluntary and written parental consent and child assent was obtained. The students accessed the virtual world in a simulated class setting. Each student was provided with a laptop computer with headphones and asked to complete each of the four 35 minute 8^{th} grade computer lessons individually. At the end of each lesson each student completed feedback questionnaires. At the end of the final grade lesson each student also completed overall assessment of the program. Sessions were observed by study personnel who logged problems (technical or content related) and provided assistance as required.

H. *Data Collection*

Questionnaire, computer-based, and observational data were collected. Demographic data included age, race, gender, computer experience, type and frequency of computer use. Attitudes to the use of computer-assisted instruction were collected using a validated questionnaire. [1, 3] Usability parameters including ease of use, understandability, acceptability, and motivation were assessed using Likert scale ratings adapted from usability assessment instruments reported by Shegog et al. (2007) [2]. Open-ended responses on recommendations for improvement of the program were collected via computer and paper and pencil questionnaires.

III. Results

In the beta test classroom environment, the study required one laptop computer for each of up to 19 students. Each laptop was pre-configured with a virtual world installation, Adobe Flash Player ™ [11], and appropriate video codecs. One laptop configured as a Windows server was deployed in the classroom for facilitator use. Scripts executed at Windows startup created the necessary network connections between students' machines and the facilitator server to enable the collection of students' responses for post-session discussion. Student rosters were loaded pre-deployment to the virtual world's database. Students participated in isolated, single-player virtual world sessions as they moved between activity stations.

Before starting a session, each student selected their name and initial activity from a pre-game menu. After a brief overview describing navigation techniques with the virtual world and session topic, students were instructed to enter the virtual world. After clicking a 2D "Go" button from the pre-game menu, students were moved to the virtual world entrance, and as students proceeded into the world, entry gates opened to reveal the first activity station (Figure 1).

Fig. 1. Virtual world entrance

A. *Demographics*

The 19 students were experienced computer users. Most accessed computers at their home (68%), at the home of friends (26%), and at school (26%). Sixty-three percent used computers for more than 1 hour on week days and 47% for more than one hour on weekends. Most frequent uses of computers were listening to music, school work, visiting websites, and e-mail (all > 50%). Approximately 63% of the sample reported achieving A or B grades at school.

B. *Usability*

Usability parameters were highly rated across and 8^{th} grade lessons. *Ease of use*: Virtual world and educational activities were rated as easy to use by 63 - 100% of students. *Acceptability*: There was 57.8 – 68.4% agreement that each of the program lessons were as much or more fun than other lessons or favorite video games. Open ended responses suggested satisfaction with the lessons with just two participants suggesting a desire for more media elements in the form of Flash-based activities and movies.

Students maneuvered in first-person point of view within the virtual world using mouse and keyboard. First person camera point of view was controlled via mouse while arrow keys controlled movement through the virtual world along the point of view axis.

Beta test results indicated that navigation cues were necessary to reduce the amount of time students used to randomly explore the virtual world during a session.

Navigation cues consisting of light paths and glowing, animated props were therefore implemented to lead students to each activity station. While students' movement was not restricted as they travelled between activity stations, inclusion of the navigation cues increased students' focus and reduced time between activities.

IV. Discussion

Cost-effective creation and deployment of a virtual world as an e-learning tool presented challenges at every stage of development. Given the scarcity of non-commercial 3D e-learning systems deployed to classrooms, the high cost associated with a commercial-grade game development engine is evident. Working around the functional deficiencies of a lower-cost, extensible game engine by incorporating in-house developed modules enables a viable solution to be deployed to the classroom within cost constraints. Yet another vital factor when considering game engine selection includes acquiring skill sets necessary to implement a viable virtual world using the engine. This includes programming as well as 3D modeling abilities, among others. These skills did not exist in-house before developing the system, and costs for such skill sets in the commercial market were found to be prohibitive. The project demonstrated that this programming effort could be achieved within the timeline with a single dedicated programmer who was not an experienced gaming programmer at the outset but who did have a robust repertoire of programming experience in data base and business applications. Overall costs for the computer-based study components included in-house development of the 3D virtual world with 2D integration totaling less than $100,000, and outsourced Flash and video production totaling less than $250,000. Costs for 3D development tools totaled less than $3,000. Lack of economic analyses in this study limits conclusions about the cost-benefit and cost-effectiveness of the program.

While much of students' interactive learning in the curriculum occurred within the 2D activities, the 3D gaming architecture enhanced the degree of student immersion within the program to the degree that the curriculum compared favorably with student's favorite video games. Using the approach presented here may present limitations to the wide-spread deployment of such systems. The targeted hardware and software configuration for the 3D virtual e-learning system was controlled and well-understood. It is unknown if an uncontrolled, general dissemination of the system to classrooms would be successful. Dissemination research efforts are underway to explore feasible dissemination strategies for this type of application within school environments.

In a constantly shifting software and technology marketplace, availability of off-the-shelf options is continually changing. While this could represent a limitation of the current work in terms of generalizing to other applications or research domains, the protocol described offers a standard approach that can be applied with other product options, thus extending its utility not only to low-cost e-learning systems but to a wide range of research and educational applications.

V. Conclusion

The protocol described above demonstrated success in developing a successful, integrated 3D/2D LAN-based educational curriculum for use in middle school classrooms that falls within limited research budget and resource constraints. Such a model has the potential to bring commercial 3D-educational interfaces within the auspices of educational researchers and curriculum developers.

References

[1] Tortolero S, Markham C, Peskin M, Shegog R, Addy R, Escobar-Chaves L, Baumler E. "It's Your Game: A Multimedia Curriculum to Prevent HIV/STI and Pregnancy for Middle Schoolers", *Under Review.*

[2] Shegog R, Markham C, Peskin M, Dancel T, Coton C, Tortolero S. "It's Your Game: An Innovative Multimedia Virtual World to Prevent HIV/STI and Pregnancy in Middle School Youth", MedInfo Congress Series Proceedings, August, 2007

[3] Markham, C., Thiel, M., Spooner, K., McKirahan, N., Shegog, R., Peskin, M., Tortolero, S., Reininger, B., Robin, L. "Using Intervention Mapping to develop, implement, and evaluate comparable abstinence education and comprehensive sexuality curricula." 135th Annual Meeting of the American Public Health Association, Washington, DC, November 3-7, 2007.

[4] Adobe Systems, Inc. (2008). Flash Professional home page. November 21, 2008. <http://www.adobe.com/products/flash/>.

[5] Conitec Datasystems, Inc. (2008). Gamestudio home page. November 2, 2008. <http://www.3dgamestudio.com/>.

[6] Google, Inc. (2007). SketchUp home page. November 19, 2008. <http://sketchup.google.com/>.

[7] Chumbalum Soft. (2008). MilkShape 3D home page. October 30, 2008. <http://chumbalum.swissquake.ch/>.

[8] Right Hemisphere. (2008). Deep Exploration Standard home page. November 20, 2008. <http://www.righthemisphere.com/products/dexp/de_std.html>.

[9] Ultimate Unwrap 3D. (2008). Ultimate Unwrap 3D home page. November 1, 2008. <http://www.unwrap3d.com/>.

[10] Adobe Systems, Inc. (2008). Photoshop home page. November 21, 2008. <http://www.adobe.com/products/photoshop/family/>.

[11] Adobe Systems, Inc. (2008). Flash Player home page. November 21, 2008. <http://www.adobe.com/products/flashplayer/>.

Acknowledgement

This study was conducted with funding from the Centers for Disease Control and Prevention under cooperative agreement SIP 4-04, "Evaluation of Abstinence-Only and Abstinence-Plus HIV, STI and Pregnancy Prevention for Middle School Student."

Three Leveled Fuzzy System for Traffic Light and Urban Traffic Control Based on Cellular Automata

Moein Shakeri[1], Arash Deldari[2], Hossein Deldari[1], Ghamarnaz Tadayon[3]
[1]Department of Computer Engineering, Ferdowsi University of Mashhad, Iran
[2]Islamic Azad University, Mashhad Branch, Iran
[3]Islamic Azad University, Science and Research Branch, Iran
mo_sh88@stu-mail.um.ac.ir, adeldari@yahoo.com, hd@um.ac.ir, tadayon@mshdiau.ac.ir

Abstract- **It's too years that computational model of cellular automata has been proposed for studying natural phenomena of world; including communication, computation, growing and development, reproducing, contesting and evolution. Vehicular travel which demands on the concurrent operations and parallel activities is increasing throughout the world, particularly in large urban areas. In this paper, to control urban traffic, we study the simulation and optimization of traffic light controllers in a city and present an adaptive fuzzy algorithm based on cellular automata properties. We have used CA for simulating transition function of density of vehicles. In the models that have been proposed till now, environmental factors like priority of intersection streets, width and length of streets and so on, have been assumed equal and therefore they have no role in making decision for changing the status of traffic light, whereas parameters like time during the entire day, density of the vehicles of the street, number of shopping centers, offices, malls,… that have plenty of returnees, have determinant effects on amount of traffic of streets. Considering mentioned notes, we have proposed a novel system that outperforms other available models. Our system has three levels; at the first level, priority of each street is computed momently, based on fuzzy rules and regarding to environmental factors. At the second level, real velocity of vehicles of every street is calculated at specific moment and eventually at the third level, by taking into account two parameters, priority of the street and amount of density behind the traffic light, decision for changing status of traffic light is done. Simulation results of our method have been shown and compared with best algorithms of two most famous available traffic light control approaches -Global and adaptive strategies-.**

I. INTRODUCTION

In the recent years there were strong attempts to develop a theoretical framework of traffic science among the physics community. Consequentially, a nearly completed description of highway traffic [7, 11], e.g., the "Three Phase Traffic" theory, was developed. This describes the different traffic states occurring on highways as well as the transitions among them. Also the concepts for modeling vehicular traffic are well developed. Most of the models introduced in the recent years are formulated using the language of cellular automata (CA) [3,4,12,14-17]. Unfortunately, no comparable framework for the description of traffic states in city networks is present. In contrast to the highway networks, where individual highway segments can be treated separated, the structure elements of city networks exert an immense influence onto the traffic dynamics [7].

In existent urban traffic systems priority of intersection streets and width and length of streets are assumed equal and therefore they have no role in making decision for changing the status of traffic light. To overcome these limitations, we proposed a novel fuzzy system that momently computes priority of the street based on fuzzy rules and regarding to environmental factors, also length of streets are not considered the same. Subsequently a cellular automata is employed for modeling density transmission and also type of movement of vehicles.

II. CELLULAR AUTOMATA

It's too years that computational model of cellular automata has been proposed to study different fields natural phenomena [1]. CA has been shown capable of yielding discrete approximations to the solutions of systems of differential equations, in terms of which much of the macroscopic physics of our world can be expressed. CA models have been applied to fluid dynamics, growth of dendritic crystals, economics, two-directional traffic flow, image processing and pattern recognition, parallel processing, random number generation, and have even been used as a model for the evolution of spiral galaxies [13].

CA is a class of discrete dynamical systems, consisting of an array of nodes (cells) of some dimension, n. Each cell can be in one of k different states at a given tick of the clock. At each discrete tick of the clock, each cell may change its state, in a way determined by the transition rules of the particular CA. The transition rules describe precisely how a given cell should change states, depending on its current state and the states of its neighbors [5]. Let n be the dimension of the lattice, k the number of states, T the transition rule function, $C_t(i_1, \dots, i_n)$ the state of the cell at position $(i_1, \dots, i_n)$ at time t, $N_t(i_1, \dots, i_n)$ the values (given in a specific order) of the neighboring cells to this location at time t. Then the

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_86, © Springer Science+Business Media B.V. 2010

dynamics of the CA is completely specified by the initial states of all the cells, C_0 , along with the recursion rule (Equation 1).

$$C_{t+1}(i_1,\cdots,i_n) = T(N_t(i_1,\cdots,i_n)) \tag{1}$$

III. RELATED WORK

Traffic dynamics bare resemblance with, for example, the dynamics of fluids and those of sand in a pipe. Different approaches to modeling traffic flow can be used to explain phenomena specific to traffic, like the spontaneous formation of traffic jams. There are two common approaches for modeling traffic [6]; macroscopic and microscopic models.

A. *Macroscopic Traffic Models:* Macroscopic traffic models are based on gas-kinetic models and use equations relating traffic density to velocity. These equations can be extended with terms for build-up and relaxation of pressure to account for phenomena like stop-and-go traffic and spontaneous congestions.

B. *Microscopic Traffic Models:* In contrast to macroscopic models, microscopic traffic models offer a way of simulating various driver behaviors. A microscopic model consists of an infrastructure that is occupied by a set of vehicles. Each vehicle interacts with its environment according to its own rules. Depending on these rules, different kinds of behavior emerge when groups of vehicles interact.

Existent traffic prevention approaches can be categorized as follows [10]:

A. *Global Strategies:* The considered global strategies are a "synchronized strategy", a "green wave strategy" and a "random strategy". In the case of the synchronized strategy all traffic lights switch synchronously to green (red) for the east (north) bound vehicles and vice versa. In the case of a green wave strategy, i.e., adjacent traffic lights switch with a defined offset. Additionally, an appropriate offset has to be determined for the green wave strategy; this is equal to the free flow travel time for the depicted case. A further investigated candidate among the global strategies is the random strategy, i.e., adjacent traffic lights switch with a random offset. Also here a quite good result is achieved [7, 8].

B. *Adaptive Strategies:* In the following three different adaptive strategies are presented. The first investigated adaptive strategy is the "switching based on the queue length". Here a traffic signal switches if the length of a vehicle queue in front of a red light trespasses a certain value. Further investigated adaptive strategies are the "switching based on waiting time" and the "switching in analogy to a neural network". In the first case a traffic light switches to red if the green phase is not used by a vehicle for a certain time. In the case of switching in analogy to a neural network the traffic lights act like integrate-and-fire neurons. More precisely the number of passed vehicles is integrated and determines the cycle time (potential) of a traffic light. After the switching process (fire) the potential is reset to zero again. The switching in analogy to a neural network strategy leads to similar results like the switching based on the waiting time [7, 8].

IV. THREE LEVELED FUZZY PROPOSED METHOD

In the proposed models that employ adjusting traffic light for urban traffic control, priority of streets of intersection and width and length of streets have been assumed equal. whereas in urban traffic parameters like time during the entire day, density of the vehicles of the street, number of shopping centers, offices, malls that have plenty of clients, have determinant effects on amount of traffic of streets and priority of each street is defined by considering these environmental factors. Taking into account these parameters we proposed a fuzzy system that priority of each street is computed at each moment, also this system doesn't set any constraint on width and length of streets. The proposed system has three levels; at the first level priority of each street is computed momently based on fuzzy rules regarding to environmental factors and features of the street. At the second level real velocity of vehicles of every street is calculated at specific moments based on factors like priority, density and standard velocity (maximum allowed velocity that's defined for each street) of the street. Then cellular automata model is applied to specify density transmission and also type of movement of vehicles in cellular spaces of each street. And finally at the third level by considering two parameters, priority of the street and amount of density behind the traffic light decision is done for changing status of traffic light and also the specific moment that status of traffic light is changed.

A. First Level of the Proposed System

In this paper, pausing coefficient is used as a significant factor to determine priority of each street. Streets with larger quantity of stopped or paused vehicles have more priority. Pausing coefficient is classified in three main categories:

- *Health and Commercial Centers:*

In these centers, density of clients is computed momently. Since number of clients directly affects on number of paused or stopped vehicles in the street, priority of street is dependant to amount of density of clients with a linear function. So fuzzy sets of health and commercial centers are defined as follows (Figure 1):

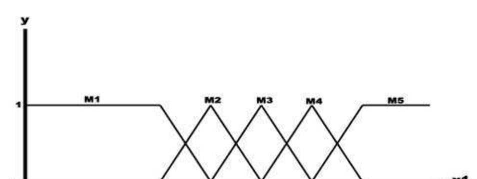


Fig. 1. Fuzzy Sets of Density of Clients of Health and Commercial Centers

In figure.1; M1, M2, M3, M4, M5 fuzzy sets are respectively related to low, high, medium, high and low densities.

- *Official Centers:*

In these centers density of clients in each moment could affect directly on priority of the street. Fuzzy sets of official centers are determined as follows (Figure 2):

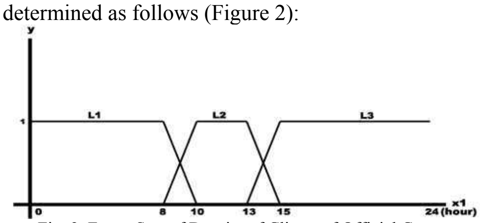


Fig. 2. Fuzzy Sets of Density of Clients of Official Centers

In figure 2; L1, L2, L3 fuzzy sets are respectively related to low, high and low densities.

- *Shopping Centers:*

Like two previous centers amount of clients of these centers, has an important effect on priority of the street, fuzzy sets of these centers can be considered as figure 3:
In figure 3; K1, K2, K3, K4, K5 fuzzy sets are respectively related to low, high, medium, high and low densities.

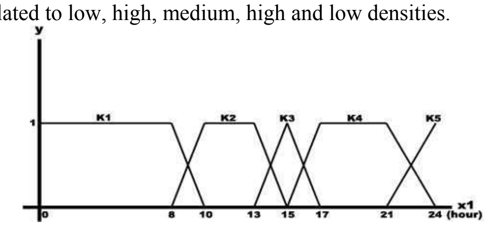


Fig. 3. Fuzzy Sets of Density of Clients of Shopping Centers

1. Fuzzy Rules

In our system, 24 fuzzy rules are regarded as follows [2]:

IF (x_1 is M_1) and (x_1 is L_1) and (x_1 is K_1) then
$F_1=[A_1*(1-y_{11})]*[A_2*(1-y_{21})]*[A_3*(1-y_{31})]$ (2)

IF (x_1 is M_1) and (x_1 is L_1) and (x_1 is K_2) then
$F_2=[A_1*(1-y_{11})]*[A_2*(1-y_{21})]*[A_3*y_{32}]$ (3)

IF (x_1 is M_1) and (x_1 is L_2) and (x_1 is K_1) then
$F_3=[A_1*(1-y_{11})]*[A_2*y_{22}]*[A_3*(1-y_{31})]$ (4)

IF (x_1 is M_1) and (x_1 is L_2) and (x_1 is K_2) then
$F_4=[A_1*(1-y_{11})]*[A_2*y_{22}]*[A_3*y_{32}]$ (5)

IF (x_1 is M_2) and (x_1 is L_1) and (x_1 is K_1) then
$F_5=[A_1*y_{12}]*[A_2*(1-y_{21})]*[A_3*(1-y_{31})]$ (6)

IF (x_1 is M_2) and (x_1 is L_1) and (x_1 is K_2) then
$F_6=[A_1*y_{12}]*[A_2*(1-y_{21})]*[A_3*y_{32}]$ (7)

IF (x_1 is M_2) and (x_1 is L_1) and (x_1 is K_3) then
$F_7=[A_1*y_{12}]*[A_2*(1-y_{21})]*[A_3*(1-|1-y_{33}|)/2]$ (8)

IF (x_1 is M_2) and (x_1 is L_2) and (x_1 is K_1) then
$F_8=[A_1*y_{12}]*[A_2*y_{22}]*[A_3*(1-y_{31})]$ (9)

IF (x_1 is M_2) and (x_1 is L_2) and (x_1 is K_2) then
$F_9=[A_1*y_{12}]*[A_2*y_{22}]*[A_3*y_{32}]$ (10)

IF (x_1 is M_2) and (x_1 is L_2) and (x_1 is K_3) then
$F_{10}=[A_1*y_{12}]*[A_2*y_{22}]*[A_3*(1-|1-y_{33}|)/2]$ (11)

IF (x_1 is M_3) and (x_1 is L_2) and (x_1 is K_2) then
$F_{11}=[A_1*(1-|1-y_{13}|)/2]*[A_2*y_{22}]*[A_3*y_{32}]$ (12)

IF (x_1 is M_3) and (x_1 is L_2) and (x_1 is K_3) then
$F_{12}=[A_1*(1-|1-y_{13}|)/2]*[A_2*y_{22}]*[A_3*(1-|1-y_{33}|)/2]$ (13)

IF (x_1 is M_3) and (x_1 is L_2) and (x_1 is K_4) then
$F_{13}=[A_1*(1-|1-y_{13}|)/2]*[A_2*y_{22}]*[A_3*y_{34}]$ (14)

IF (x_1 is M_3) and (x_1 is L_3) and (x_1 is K_2) then
$F_{14}=[A_1*(1-|1-y_{13}|)/2]*[A_2*(1-y_{23})]*[A_3*y_{32}]$ (15)

IF (x_1 is M_3) and (x_1 is L_3) and (x_1 is K_3) then
$F_{15}=[A_1*(1-|1-y_{13}|)/2]*[A_2*(1-y_{23})]*[A_3*(1-|1-y_{33}|)/2]$ (16)

IF (x_1 is M_3) and (x_1 is L_3) and (x_1 is K_4) then
$F_{16}=[A_1*(1-|1-y_{13}|)/2]* [A_2*(1-y_{23})]*[A_3*y_{34}]$ (17)

IF (x_1 is M_4) and (x_1 is L_2) and (x_1 is K_2) then
$F_{17}=[A_1*y_{14}]*[A_2*y_{22}]*[A_3*y_{32}]$ (18)

IF (x_1 is M_4) and (x_1 is L_2) and (x_1 is K_3) then
$F_{18}=[A_1*y_{14}]*[A_2*y_{22}]*[A_3*(1-|1-y_{33}|)/2]$ (19)

IF (x_1 is M_4) and (x_1 is L_2) and (x_1 is K_4) then
$F_{19}=[A_1*y_{14}]*[A_2*y_{22}]*[A_3*y_{34}]$ (20)

IF (x_1 is M_4) and (x_1 is L_3) and (x_1 is K_2) then
$F_{20}=[A_1*y_{14}]*[A_2*(1-y_{23})]*[A_3*y_{32}]$ (21)

IF (x_1 is M_4) and (x_1 is L_3) and (x_1 is K_3) then
$F_{21}=[A_1*y_{14}]*[A_2*(1-y_{23})]*[A_3*(1-|1-y_{33}|)/2]$ (22)

IF (x_1 is M_4) and (x_1 is L_3) and (x_1 is K_4) then
$F_{22}=[A_1*y_{14}]*[A_2*(1-y_{23})]*[A_3*y_{34}]$ (23)

IF (x_1 is M_5) and (x_1 is L_3) and (x_1 is K_4) then
$F_{23}=[A_1*(1-y_{15})]*[A_2*(1-y_{23})]*[A_3*y_{34}]$ (24)

IF (x_1 is M_5) and (x_1 is L_3) and (x_1 is K_5) then
$F_{24}=[A_1*(1-y_{15})]*[A_2*(1-y_{23})]*[A_3*(1-y_{35})]$ (25)

In the above rules y_{1j},y_{2j} and y_{3j} respectively demonstrate membership functions of M_j,L_j and K_j fuzzy sets.

2. Defuzzification

In this system defuzzification is performed by using equation (26). Output of defuzzifier is considered as output of first level of system and shows priority of the street.

$$F = \frac{\sum_{i=1}^{24} Fi}{n} \tag{26}$$

In the above equation, Fi is output of ith rule (i=1...24), n is number of rules that have nonzero output and F is defuzzifier output (output of first level of system). F uses as one of the inputs of second level of the system.

B. Second Level of the Proposed System

In this step real velocity of vehicles of every street is calculated at specific moments and then type of transmission of density of vehicles is evaluated. Real velocity of vehicles is obtained by regarding three factors:

- *Priority of Street:* That's calculated dynamically at specific moments, based on fuzzy rules at the first level.
- *Density of Vehicles:* Since low-priority streets have low pausing coefficients, most of vehicles would be paused behind intersections. So if adjustment of the traffic light is done only based on priority of the street, vehicles may pause for a long duration behind intersections. To prevent this problem, number of existent vehicles in the street is also regarded as one of the inputs.
- *Standard Velocity of Vehicles of Street:* Since number of vehicles behind intersections, is a determinant factor for specifying distance between pausing location and traffic light, we have used cellular automata for segmentation of street from intersection to end of street. So type of movement of vehicles from entrance time till pausing time behind intersection is achieved by using transition rules of cellular automata. standard velocity of vehicles (standard velocity is constant for all vehicles of each street and computed based on width and length of street) must be specific while applying transitions rules so it's considered as third factor.

Based on these three inputs, velocity of each vehicle is computed as follows:

$$V=(1-\gamma)(1-\alpha)*V_N+\varepsilon \tag{27}$$

γ is priority of the street (output of first level of fuzzy system), V_N is standard velocity of street, ε is a constant value for movement of vehicles (5km/h in this paper) and α is density of the street that's obtained from equation 28 :

$$\alpha = \frac{N}{W*L*A} \tag{28}$$

W and *L* are respectively width and length of the street, *A* is the average occupied area by each vehicle (15 m2 in this paper), *N* is number of available vehicles in the street (*N* is calculated as follows: *W*L* is multiplied by summation of densities of cellular spaces). Regarding equation 28 more density or more priority of street, leads in less velocity of vehicles. Based in this velocity, transition rules of cellular automata are determined.

1. Cellular Automata Rules

Since velocity and displacement have a linear dependency (x=v*t), we used a linear transition function. Because of extra computations, discrete time method is applied instead of continuous time approach and system is updated at specific seconds. In this paper, velocities of vehicles are divided in intervals of length 5 and length of each cell is considered as 10 meters. So at every 7.2" time interval, the vehicle moves 10 meters (equal to one cell) in a fixed 5km length. (Updating progress is done every 7.2 seconds). Based on these comments, we used following transition function for obtaining cellular rules of density transmission in the street:

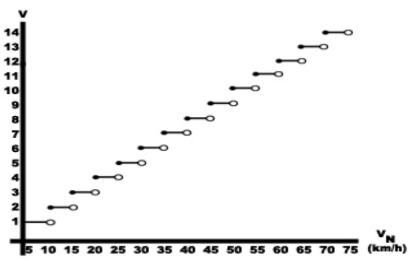


Fig. 4. Density Transmission Function Based on CA Model

In the proposed cellular automata model width of cell is regarded as width of the street and length of cell is set to a constant number (10). Note that transmission of density of vehicles means transmission of vehicles from one cell to another one, type of this transmission is specified as figure 4.

Example: if initial velocity of vehicle is 45 km/h, equation 28 is used for transmission of density:

$$C(j+9)=C(j+9)+C(j) \tag{29}$$

Where C(j+9) is 9th neighbor of C(j) in the direction of movement. Maximum density of each cell is one and related to number of available vehicles at each cell. Maximum number of vehicles of each cell is computed as follows:

$$N = \frac{w*10}{15} \tag{30}$$

W is width of the street, 10 is length of the street and 15 shows estimated area allocated for a vehicle.

If number of available vehicles of each cell exceeds the maximum value (calculated from equation 30) surplus is transferred to previous cell. Notice that in this paper all of the streets are bidirectional and cellular space of streets is as follows:

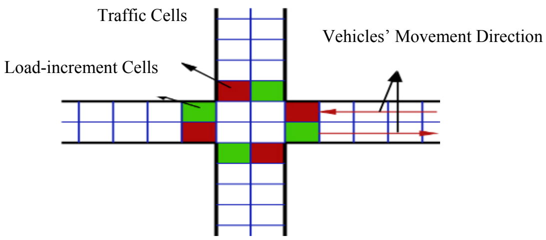


Fig. 5. Cellular Space of Streets and Corresponding Traffic and Load-increment Cells

Traffic cells that illustrated in figure 5 have different affects, when the traffic light is red or yellow these cells cause to increment of cell density, whereas they cause to decrement of cell density when the traffic light is green. Load-increment cells always cause to transmission of density from one street to another one.

In this paper transmission of density from traffic cells to load increment cells is done equally and steady. Traffic cells of the street momently dispread the density equally among load-increment cells of three other streets. Dispreading of density also can be done regarding to type of the street.

C. Third Level of the Proposed System

In this paper two factors are considered for adjusting traffic light. Priority of the street and number of complete traffic cells are inputs of third level of the proposed system. Since density and traffic in high-priority streets have more undesirable effects, priority is considered for adjusting traffic light. Furthermore number of paused vehicles behind intersections is the main factor to determining time for changing status of traffic light. This factor is illustrated by traffic cells and regarded as input of system.

In this proposed system two conversion functions are used for priority and number of traffic cells to determine time duration that green light is active.

Multiplication of outputs of two functions represents time duration that traffic light is green for each street:

$$G=Tc*\gamma \tag{31}$$

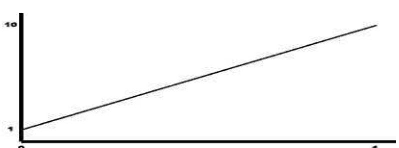


Fig. 6. Conversion Function (Priority to Time)

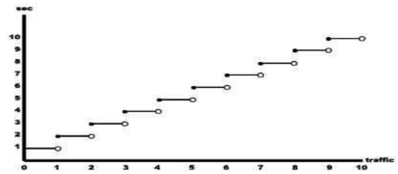


Fig. 7. Conversion Function (Number of Traffic Cells to Time)

G shows the time interval that green light is active, *Tc* is number of traffic cells and γ is priority of the street. *G* is computed for all of streets of intersection. These calculations are done concurrently and independently and finally street with the highest *G* actives green light and uses *G* seconds from the green light. Like section 3-2-1 the system is updated every 7.2 seconds. To prevent consecutive variations of traffic light status equation 32 is employed:

If ***G_max >= 2*G_green_light*** THEN (32)
Change ***Traffic light***

G_max is maximum value of *G* in every updating progress (among streets of an intersection) and *G_green_light* is updated value of G for the street with green light. *Change Traffic Light* instruction actives green light for the street that has G_max. In the proposed approach traffic every moment light green is active only for one of the streets.

V. SIMULATION RESULTS

The proposed method was simulated with specific inputs and more desirable results were obtained in comparison with based adaptive method.

When priorities of streets are considered almost equal and factors like average velocity of street and density are disregarded, three level fuzzy and NN-based approaches almost have same responses.

When priorities of streets are different from each other and density of vehicles of streets of intersection streets is not the same results of NN-based method doesn't change at all, because this method doesn't consider priority of streets, so density of high-priority streets would increase that this matter leads to traffic increment of that street. Whereas in the proposed three level fuzzy approach changing priority of streets and density of each street causes to significant changes in traffic light scheduling and traffic cells. This change results in only street with more density and higher priority adjust traffic light. Finally traffic load and street density will be balanced.

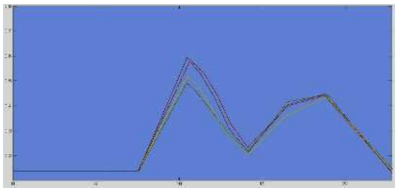


Fig. 8. Density of Streets of Intersection Obtained from Fuzzy Proposed Approach (Blue Shows High-priority and Green Show Low-priority Street)

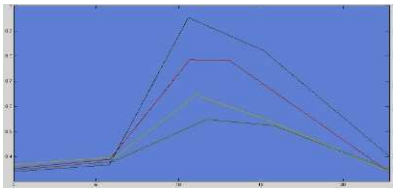


Fig. 9. Density of Streets of Intersection Obtained from NN-based Approach (Blue Shows High-priority and Green Show Low-priority Street)

Figure 8 indicates two main notes, First; the proposed system controls density of intersecting streets in a way that traffic load would be balanced and equal and second; higher priority of streets leads to more density at intersections.

For better understanding we review an example: we consider intersection's streets have these main centers:

- 1st street: 5 Health and Commercial Centers, 4 Official Centers, 1 Shopping Center
- 2nd street: 4 Health and Commercial Centers, 5 Official Centers, 1 Shopping Center
- 3rd street: 2 Health and Commercial Centers, 3 Official Centers, 1 Shopping Center
- 4th street: 1 Health and Commercial Centers, 1 Official Centers, 2 Shopping Center

Note that priorities of streets would be as follows (based on fuzzy relationships):

- 1st street: 0.8333 - 2nd street: 0.8333
- 3rd street: 0.25 - 4th street: 0.0833

It's assumed that standard velocity of all streets are the same and equal to 50 km/h, also width and length of streets are considered as 10m and 2km. so real velocity would be:

- 1st street: 13.3194 - 2nd street: 13.3194
- 3rd street: 42.4375 - 4th street: 50.5769

At third level of the system density and number of traffic cells at 12 o'clock are computed as follows:

- 1st street:Density: 0.5834,Number of Traffic Cells: 2
- 2nd street:Density:0.5741,Number of Traffic Cells: 2
- 3rd street:Density:0.5208,Number of Traffic Cells: 3
- 4th street:Density: 0.4917,Number of Traffic Cells: 4

Whereas by regarding NN-based method density and number of traffic cells would be as:

- 1st street:Density: 0.9571,Number of Traffic Cells: 2
- 2nd street:Density:0.7840,Number of Traffic Cells: 2
- 3rd street:Density:0.6425,Number of Traffic Cells: 3
- 4th street:Density: 0.5403,Number of Traffic Cells: 4

Notice that although inconsiderable density has been added to 4th street, density decrement is completely sensible in 1st,2nd and 3rd streets.

While existent systems only consider specific limited segment behind the intersection for adjusting traffic light, our proposed system can control density of vehicles and traffic load throughout the street and prevent heavy traffics.

VI. CONCLUSIONS

In this paper a three level fuzzy system was proposed for urban traffic control by adjusting traffic light. Using this approach results in traffic decrement and also steady traffic dispreading all over the city. In the proposed system priority and length of streets are considered variable. Also real velocity of vehicles is computed based on fuzzy rules in different hours during the day. Considering this factor leads to traffic load balancing throughout the city. Another dominant note in this paper is that priority of streets and max allowed velocity of the street are obtained by fuzzy sets with regard to urbanism principals. This matter also reduces traffic significantly.

REFRENCES

[1] Stephen Wolfram, "A New Kind of Science" , 2002

[2] Li-Xin Wang , "A Course In Fuzzy Systems and Control", 1996

[3] M.Schreckenberg, R.Barlovic,W. Knospe,H. Klupfel, "Statistical Physics of Cellular Automata Models for Traffic Flow" , Phd thesis,Germany ,2004.

[4] S. Wolfram (ed.). "Theory and applications of CA", World Scientific, Singapore, 1986

[5] Ron Breukelaar and Thomas Back, "Evolving Transition Rules for Multi Dimensional Cellular Automata", Springer-Verlag Berlin Heidelberg 2004

[6] M. Wiering , J. van Veenen , J. Vreeken , A. Koopman, "Intelligent Traffic Light Control", Institute of Information and Computing Sciences, Utrecht University, 2004

[7] R.Barlovic , T.Huisinga , A.Schadschneider , M.Schreckenberg, "Adaptive Traffic Light Control in the ChSch Model for City Traffic", Duisburg University, 2005

[8] E. Brockfeld , R. Barlovic , A.Schadschneider , M. Schreckenberg ,"Optimizing Traffic Light in a cellular automaton model for City Traffic " , Duisburg University, PPHYSICAL RREVIEW E, VOLUME 64, 056132 , 2001.

[9] J. Neal Richter, David Peak , "Fuzzy Evolutionary Cellular Automata", Proccedings of ANNIE ,2002.

[10] B.S. Kerner, " Empirical macroscopic features of spatial – temporal traffic patterns at highway bottlenecks", Physical Review, Volume 65,046138, Stuttgart, Germany,2002.

[11] Traffic and Granular Flow, '97, '99, '01, Springer, (1996,1998,2000,2003).

[12] B.S. Kerner, "Traffic behavior near an off ramp in the cellular automaton traffic model" , Phys. Rev. E 65, 046138 (2002).

[13] R. Espericueta,"Cellular Automata Dynamics", Math Department , Bakersfield College, 1997.

[14] K. Nagel and M. Schreckenberg, J. Physique I 2, 2221 (1992).

[15] N.H. Gartner and N.H.M. Wilson (eds), Transportation and Tra_c Theory, (New York: Elsevier,(1987).

[16] M. Fukui and Y. Ishibashi, J.Phys. Soc. Jpn. 62 (1993).

[17] P. Wagner, K. Nagel, and D.E. Wolf, Physica A 234, 687 (1997).

Inter-Machine Cooperation Mechanism for Dynamic Scheduling

Ana Madureira, Joaquim Santos
Institute of Engineering – Polytechnic of Porto, Portugal

Abstract— **This paper addresses dynamic scheduling resolution and suggests the interest of hybrid approaches trough MAS coordination and Meta-Heuristics. The proposed Multi-agent Scheduling System assumes the existence of several Machines Agents (which are decision-making entities) distributed inside the Manufacturing System that interact and cooperate with other agents in order to obtain optimal or near-optimal global performances. Agents have to manage their internal behaviors and their relationships with other agents through cooperative negotiation in accordance with business policies defined by the user manager. An Inter-Machine Cooperation Mechanism for Dynamic Scheduling is proposed.**

***Index Terms*—Cooperation, Dynamic scheduling; distributed agent system; meta-heuristics; manufacturing systems.**

I. Introduction

Scheduling Coordination can be seen as the intelligent allocation of resources among individuals and their goals in social society, which makes individuals using their resources more beneficially, so as to promote the achievement of individual and/or social goals. Usually, the requirement of coordination comes from the distribution and interdependencies of resources, entities and information. In scheduling coordination, considering the resource constraints, individual plans are to be evaluated and refined, and it may be needed to acquire available resources from external environment, including other individuals.

Agent-based computing is a promising approach for developing applications in complex domains, a number of challenges still need to be faced to turn agent-oriented software abstractions into practical tools for facing the complexity of a modern application like solving dynamic and distributed scheduling problems. Probably the most important challenge is to discover the best option to define the way like all the agents present in a system communicate and interact among them. Coordination of Multi-Agent Systems is an important clue when this approach is considered to solve distributed and complex problem like the one that is object of our work.

Bio-Inspired Techniques form a class of powerful and practical solution techniques for tackling complex, large-scale combinatorial problems producing efficiently high-quality solutions. From the literature we can conclude that they are adequate for static problems. However, real scheduling problems are quite dynamic, considering the arrival of new orders, orders being cancelled, machine delays or faults, etc. Scheduling problem in dynamic environments has been investigated by a number of authors [1][2].

This paper addresses dynamic scheduling resolution and suggests the interest and utility of hybrid multi-agent coordination. This work is included in project MASDScheGATS [1] that aim to continuously and efficiently adapts the current solution to a changing environment.

The remaining sections are organized as follows: section II summarizes some literature review on Multi-Agent Systems. Section III presents an overview about coordination in Multi-Agent Systems. In section IV MASDScheGATS System, the developed system, is described and a coordination algorithm is proposed in order to allow the improvements of systems' performance and reliability. Finally, the paper presents some conclusions and puts forward some ideas for future work.

II. MULTI-AGENT SYSTEMS

Multi-Agent Systems (MAS) definition depends on the domain where agents are used and because of that MAS definitions abound. We believe that a solid and generic definition is presented by [2] that define MAS as a system capable to "solve complex problems in a distributed fashion without the need for each agent to know about the whole problem being solved". This definition and many others imply the existence of an organizational model, or management mechanism, that allows for interaction, communication and achievement of objectives within the group of agents, regardless whether the system is comprised of self-interested, group oriented or both types of agents. A proper definition for such mechanism is given by the definition of the word coordination, particularly if one considers coordination applied to MAS. Notice that in [3] we can find several reasons on why agents need to be coordinated:

- Chaos and anarchy prevention – coordination is desirable because in distributed system anarchy is easy to implement,
- Global restrictions – agents must obey to a set of

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_87,

restrictions in order to be well succeeded,

• Knowledge, resources and information is distributed – agents can have different capacities and specializations. Alternatively agents can have different information resources, responsibilities and limitations.

• Dependency between agents actions – agents' objectives are normally inter-dependent, a situation that requires coordination of actions.

• Efficiency – autonomous agents can work in an independent way, but at the same time they may have the necessity to coordinate their actions with other agents. The information of an agent can be sufficient for in group with another agent solve a problem quickly.

Scheduling problems arise in a diverse set of domains, ranging from manufacturing to hospitals settings, transports, computer and space environments, amongst others. Most of these domains are characterized by a great amount of uncertainty that leads to significant system dynamism. Such dynamic scheduling is receiving increasing attention amongst both researchers and practitioners.

In spite of all previous contributions, the scheduling problem is known to be NP-complete [4]. This fact incites researchers to explore new directions and Multi-Agent technology has been considered an important approach for developing industrial distributed systems [5-6].

III. COORDINATION IN MULTI-AGENT SYSTEMS

In a real manufacturing system a product is produced, step by step, passing on several machines. In each machine it will be performed at least one operation (job) of the process plan. However, if solutions obtained are joined by machine agents will be observed that in most of cases the founded solution is not feasible. In fact, if operation Op1 (in machine m1) precedes operation Op2 (in machine m2) and Op2 precedes Op3 (in machine m3) in a manufacturing process, it is not guaranteed that the initial time for Op2 in m2 will be after the end of Op1 in m1 nor that the end of Op2 in m2 will be before the start of Op3 in m3.

Cooperation and negotiation are the two main types of coordination protocols developed in the literature and implemented in the existing systems. Therefore some of these protocols and mechanisms will be listed in the next paragraphs.

Cooperation is generally defined in the literature as the act of combining efforts, in order to accomplish a common objective that one autonomous agent alone cannot reach by itself. In other words, "cooperation refers to a coordination protocol among nondisputant agents" [7]. Such protocol is suited when agents need to share tasks or results as a way to reach the system's objective, so normally it is used in team like based architectures.

On the other hand, "negotiation is the coordination among competitive or simply self-interested agents"[7]. Negotiation can be described like the process in which at least two agents negotiate through a protocol in order to accomplish an agreement. This protocol is suitable for systems composed by a set of agents that pursuit their own individual goals but need interaction with other agents to achieve a satisfactory outcome. The buyer-seller situation is the best analogy to demonstrate this protocol. Negotiation is normally used in market like based approaches.

In [6] can be encountered a recent state of art related with the application of Multi-Agent Systems to solve production planning and scheduling problems.

There exist some important approaches on agent-based production scheduling and control [8] [9], therefore, here we concentrate on the most important aspects.

A Collaborative coordination control (CCC) mechanism is proposed in [5].

In [10] a hybrid, agent-based scheduling and control system architecture is presented in order to solve task allocation problems.

A multi-agent software system RIDER (Real-Time Decision Making in Manufacturing) has been developed for a cable producing company and for a carpet manufacturer [12]. Upon the occurrence of an event (machine breakdown, new orders, etc.), the agents use a mechanism for generating local alternatives and follow a message exchange procedure to build decision trees, which are evaluated by user defined cost-based objective functions.

A system that uses an agent-based collaborative production control framework capable of conducting scheduling and dispatching functions among production entities is reported in [13].

Cooperation and conflict resolution are the main concerns in agent-based scheduling systems. For solving this kind of problems, the idea of negotiated factory scheduling are normally considered.

A market-based negotiation mechanism, called precedence cost tâtonement (P-TÂTO), is described in [14]. The system is composed of a project manager agent, task agents, resource manager agents, resource agents, and coordinator agents.

In [15] and [16] a market based mechanism was proposed which made cooperation in the control of distributed manufacturing systems possible. The system included self-interested, autonomous agents that the main objective is to pursuit their goal. The planning problem is responsibility of a central agent, machine agents were responsible for generate and execute local generated schedules at machines. Non feasible schedules was the responsibility of the central agent, but scheduling decisions were reached using a system of bides between the central agent and the machine agents.

Stimergy is a coordination mechanism for multi-agent systems based on animal interactions i.e. animal-animal interactions, evolved significantly and, presently, the analogy with the social insects confuses rather than helps the discussion [17]. This system is a coordination protocol applied to PROSA reference architecture [18]. It is inspired by natural systems i.e. food foraging behavior in ant colonies. Stigmergy is an indirect coordination tool within an insect society where parts of global information is made available locally by pheromones informing other ants about remote facts.

In [19], the problem of distributed resource allocation in collaborative environment, such as distributed manufacturing environment or network of enterprise in supply chain, is a main focus specially in the environment where there is a high degree of autonomy among the independent entities, e.g. factories, companies, and enterprises. It is based in a centralized control, where the given tasks are assigned based on a specific resource allocation algorithm, the coordination protocol guides the interaction among the independent entities (task and resource owners) via communication channel.

Xue [20] introduces a system where two scheduling strategies are used to optimize the solution, the first earliest-delivery-time-based that allow to give the product the customer as soon as possible and the second due-time-based to start the product manufacturing as late as possible in order to reduce the time to store the finished product. Constraint based search and agent-based are used to identify the optimal schedule.

In [21] the allocation of resources to production jobs is carried out in a way similar to an open marketplace. Each part agent tries to full the processing requirements to achieve its own set of objectives (due date, quality, costs, etc.); for this purpose it is given a certain amount of currency, representing its own purchase capacity, in order to acquire the needed service from the single production resources.

Tseng et al [22] proposed a collaborative control structure is proposed for the mass customization manufacturing system. Each workstation is considered as an autonomous agent seeking the best return. The individual work order is considered as a job agent that vies for the lowest cost for resource consumption. The system schedule and control are integrated as an auction based bidding process with a price mechanism that rewards product similarity and response to customers' needs.

A MAS-based platform [23] to integrate manufacturing planning and control with systems reconfiguration and restructuring is proposed. This system has the capability to model complex heterogeneous systems, their structures and physical constraints, and also product orders. It allows the hierarchical structures of complex manufacturing systems to be modeled but avoids centralized control in classical hierarchical and hybrid control frameworks.

In [24] are developed some extension to the work represented in [20] in order to introduce a reactive mechanism for responding to changes of production orders and manufacturing resources.

Dumond et al [25] argues that algebra provides an appropriate context both for specifying the agents and their global coordination. They use the p- calculus for the description of the following entities, the internal behavior of the agents, i.e. the treatments to be achieved in response to external requirements, the agents' respective interfaces with the other agents, the coordination layer, i.e. the global outline of the agents' collaboration.

Shehory et al [26] adopt methods used by physicists to study interactions among multiple particles. The physics-oriented methods are used to construct a coordinated task-allocation algorithm for cooperative goal-satisfaction.

In [27] there are enumerated very important approaches referents to agent coordination like mutual adjustment, direct supervision, coordination by standardization, mediated coordination, coordination by reactive behaviour.

IV. MASDSCHEGATS SYSTEM

Distributed environment approaches are important in order to improve scheduling systems flexibility and capacity to react to unpredictable events. It is accepted that new generations of manufacturing facilities, with increasing specialization and integration, add more problematic challenges to scheduling systems. For that reason, issues like robustness, regeneration capacities and efficiency are currently critical elements in the design of manufacturing scheduling system and encouraged the development of new architectures and solutions, leveraging the MAS research results.

The work reported in this chapter is concerned with the resolution of realistic scheduling problems, called here Extended Job-Shop Scheduling Problems (EJSSP) [1]. Moreover, it is concerned with integrated scheduling of jobs which are products composed by several parts or components that may be submitted to a number of manufacturing and multi-level assembly operations.

The EJSSP scheduling problem is decomposed into a series of Single Machine Scheduling Problems (SMSP) which are independently solved and results incorporated into the main EJSSP problem. MASDScheGATS (Multi-Agent System for Distributed Manufacturing Scheduling with Genetic Algorithms and Tabu Search) system, which is a Team-Work based MAS, assumes the existence of several machines agents that interact and cooperate with other agents in order to obtain optimal or near-optimal global performances. The idea is that from local, autonomous and often conflicting agent's behaviors a global solution emerges from a community of machine agents solving locally their schedules trough cooperative negotiation with other machine agents in order to increase schedule quality.

Machine Agents (which has a Meta-Heuristic associated) represents a SMSP and it is responsible for scheduling this resource. We can name Machine Agents like Meta-heuristic Agents, considering that each Machine Agent performs a Meta-heuristic Algorithm. As the Machine agents find their respective local optimal, they must negotiate with other agents in order to overcome inter-agent constraints and achieve a global schedule.

The work described in this paper is a system where a community of distributed, autonomous, cooperating and asynchronously communicating machines tries to solve scheduling problems.

The main purpose of MASDScheGATS is to decompose the scheduling problem into a series of Single Machine Scheduling Problems (SMSP) and create a Multi-Agent system where each agent represents a resource (Machine Agents) in a Manufacturing System [28]. Each Machine Agent must be able:

- to find an optimal or near optimal local solution trough Genetic Algorithms or Tabu Search meta-heuristics.
- to deal with system dynamism (new jobs arriving, cancelled jobs, changing jobs attributes, etc).
- to change/adapt the parameters of the basic algorithm according to the current situation.
- to switch from one Meta-Heuristic algorithm to another.
- to cooperate with other agents.

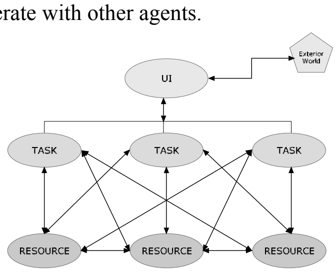


Figure. 1. MASDScheGATS System Architecture

The proposed architecture is based on three different types of agents: User Interface, Task agents and Resource Agents. In order to allow a seamless communication with the user, a User Interface Agent is implemented. This agent, apart from being responsible for the user interface, will generate the necessary Task Agents dynamically according to the number of tasks that comprise the scheduling problem and assign each task to the respective Task Agent.

The Task Agent will process the necessary information regarding the task. That is to say that this agent will be responsible for the generation of the earliest and latest processing times, the verification of feasible schedules and identification of constraint conflicts on each task and the decision on which Machine Agent is responsible for solving a specific conflict.

Finally, the Machine Agent is responsible for the scheduling of the operations that require processing in the machine supervised by the agent. This agent will implement meta-heuristic/local search procedures in order to find best possible operation schedules and will communicate those solutions to the Task Agent for later feasibility check (Figure. 1).

V. Coordination Mechanism

We propose an algorithm that in each time elects a machine to cooperate with another one. This process of election is based in the number of conflicts that each has in a determinate moment. The elect machine will be the one with the higher number of conflicts. After each negotiation round (we consider a round the act of a specific machine "negotiates" with another one) all conflicts are counted again and the machine with higher number of conflicts will be elected to cooperate. The elected machine will try to cooperate with other machines considering operations ordered by descending of processing time [28].

TABLE I
JOB DATA FOR THE EXAMPLE

Job	*Operation*	*Machine*	p_{ijkl}	t_{ijkl}
j_1	o_{1111}	m_1	5	0
	o_{2121}	m_2	7	5
	o_{3131}	m_3	4	12
j_2	o_{1211}	m_1	2	0
	o_{2221}	m_2	9	2
	o_{3231}	m_3	4	11
j_3	o_{1311}	m_1	8	0
	o_{2321}	m_2	2	8
	o_{3331}	m_3	7	10

Initialization

Each machine is initialized with a flag (flagM1, flagM2, …, flagMn) that will be used to control which machine has been elected in the last coordination round. In the beginning the all the machine flags are equal to zero, and after each round the nominated machine flag will be -1. A machine cannot be elected in two contiguous rounds.

In the beginning a penalization for each task (job) is generated based on the it delivery date. For example if the delivery date of j1 is 30, j2 is 25 and j3 is 32, the penalization will be 2 for j1 , 3 for j2 and 1 for j3 .

Each operation inherits the penalization from its job. With these values a penalization for each machine is calculated based in the penalizations sum of operations that are in conflict in each machine.

A variable count (countM1, countM2, ..., countMn) is created for each machine to save the number of time that a machine is elected to cooperate.

A. *Algorithm*

The machine with total higher number of conflicts will be elected to cooperate with the one with it has more conflicts. In a case in which more than one machine has the higher number of conflicts with the elected one, the penalization calculated for each machine will be used and the one with higher penalization will cooperate.
When the total number of conflicts of one machine reaches zero, the cooperation mechanism will not allow any change in the scheduling plan that can affect this.

```
// Inicialization
arrayFlags[NumberofMachines]
arrayCountElection[NumberofMachines]
arrayPen[NumberofJobs]
arrayDeliveryDates[NumberofJobs]

For var=0 to var= NumberofMachines-1
// Put arrayflags positions equals to zero
Loop

For var1=0 to var1= NumberofJobs-1
// create penalizations for each job based on each
delivery date
Loop

Function verifyConflicts
For var2=0 to NumberofMachines-1
// Verify conflits for each machine.
// Discover the two machines with the higher number
of conflicts. This function verifies if the machine
with more conflicts was the last to be elected. If
it is true the second will be the elected.  The two
machines that will coordinate are returned.
Loop

End function

FunctionSolveConflicts
//  Function verifyConflicts is called. If zero
conflicts are reached processing is stopped.
// The higher one between these two will be elected
and will negotiate with the other.
//ArrayFlags and ArrayCountElections are changed.

End Function
```

B. *Illustrative Example*

Step1: Generate locally in each machine a solution using an MetaHeuristics.

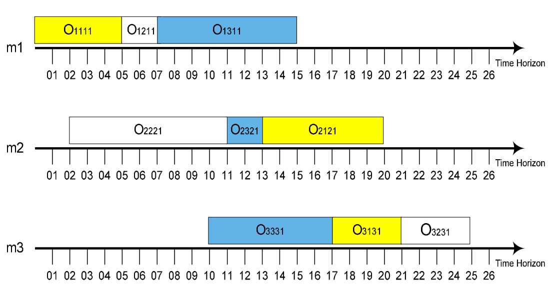


Figure 2. Schedule generated by Meta heuristics (Results of 1st step)

Machine agents will send local plans to Task agents that will recognize all conflicts. For example, we have the following conflicts: O_{1211} and O_{2221}, O_{2121} and O_{3131}, O_{1311}, O_{2321}, O_{3331}.

TABLE II
EXAMPLE OF MACHINE CONFLICTS

	M1	M2	M3
M1	0	2	1
M2	2	0	2
M3	1	2	0
Total	**3**	**4**	**3**

On machine M1 we identify conflicts on operations O_{1211} and O_{1311}. The penalization is PenM1 = 2 + 1 = 3.
On machine M3 we identify conflicts on operations O_{3331} and O_{3131}. The penalization is PenM3 = 1 + 2 = 3.

Step 2: The M2 machine will cooperate with M1. Operation cooperation order O_{2221}, O_{2121}, O_{2321}.

Results of step 2

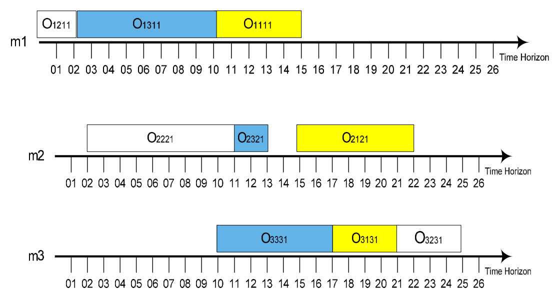


Figure 3. Schedule plan after 1st step

The flagM2 will be set to minus 1 and countM2 is increment by one.

TABLE III
EXAMPLE OF MACHINE CONFLICTS

	M1	M2	M3
M1	0	0	0
M2	0	0	2
M3	0	2	0
Total	**0**	**2**	**2**

Step 3: The M3 machine will cooperate with M2. The number of conflicts of M2 and M3 is the same, but flagM2 is set to minus 1 so M3 will be elected. Operation cooperation order O_{3331}, O_{3231}, O_{3131}.

In this case O_{3231}, O_{3131} have the same processing time so the alphabetical order is considered to create operation

Results of step 3

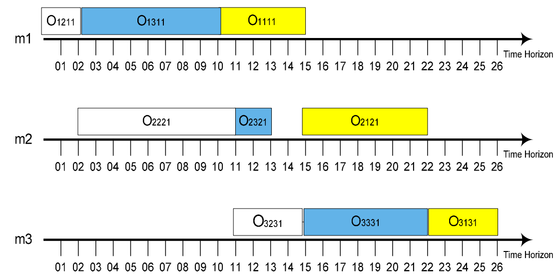


Figure 4. Final Schedule Plan

TABLE IV
EXAMPLE OF MACHINE CONFLICTS

	M1	M2	M3
M1	0	0	0
M2	0	0	0
M3	0	0	0
Total	**0**	**0**	**0**

VI. CONCLUSIONS AND FUTURE WORK

We believe that a new contribution for the resolution of more realistic scheduling problems (Extended Job Shop Problems) was described in this paper. The particularity of our approach is the procedure to schedule operations, as each machine will first find local optimal or near optimal solutions, succeeded by the interaction with other machines trough cooperation as a way to find an optimal or near-optimal global schedule.

A Coordination Protocol for Team-work based Architecture was proposed based upon a cooperation mechanism.

Work still to be done includes the testing of the proposed system and negotiation mechanisms under dynamic environments subject to several random perturbations.

ACKNOWLEDGMENT

The authors would like to acknowledge FCT, FEDER, POCTI, POCI for their support to R&D Projects and GECAD - Knowledge Engineering and Decision Support Group Unit.

REFERENCES

[1] Ana Madureira, , Joaquim Santos, , Developing a Multi-Agent System for Distributed Manufacturing, 2006, Knowledge and Decision Technologies, ISBN: 972-8688-39-3.

[2] B.I Kim, S.S. Heragu, R.J. Graves, A. St Onge, A Hybrid Scheduling and Control System Architecture for Warehouse Management, 2003, IEEE Trans. on Robotics and Autom., 19/6: 991-100

[3] H. Nwana, L. Lee, N. Jennings, Coordination in software agent systems, 1996, BT Technol J Vol 14, No 4 October,.

[4] J. Blazewicz, K. H. Ecker, E. Pesch, G. Smith, J. Weglarz, Scheduling Computer and Manufacturing processes, 2001, Springer, 2nd edition, New York.

[5] W. Shen, and D. Norrie, Agent-based systems for intelligent manufacturing: a state of the art survey, 1999, Int. J. Knowl. Inform. Syst., vol. 1, no. 2, , pp. 129– 156.

[6] M. Yokoo and K.Hirayama, Algorithms for Distributed Constraint Satisfaction: A Review, 2000, Journal of Autonomous Agents and Multi-Agent Systems,.

[7] G. Weiss, MultiAgents Systems – A Modern Approach to Distributed Artificial Intelligence, 1999, The MIT Press.

[8] M.Caridi, , Cavalieri, S., Multiagent Systems in Production Planning and Control: An Overview, Production Planning and Control, 2004, 15/2: 106-118.

[9] R.F. Babiceanu, F.F. Chen, Development and Applications of Holonic Manufacturing Systems: A Survey, 2006, Journal of Intelligent Manufacturing, 17: 111-131.

[10] S.Y. Nof, R. Weill, Collaborative Coordination Control (CCC) of Distributed Multimachine Manufacturing, 1992, Annals of the CIRP, 41/1: 441-444.

[11] B.I Kim, S.S.Heragu, R.J. Graves, A. St Onge, A Hybrid Scheduling and Control System Architecture for Warehouse Management, 2003, IEEE Trans. on Robotics and Autom., 19/6: 991-100

[12] N. Papakostas, D. Mourtzis, K. Bechrakis, , G. Chryssolouris,, D. Doukas, R. Doyle, A Flexible Agent Based Framework for Manufacturing Decision-Making, 1999, Proc. of the 9th Flexible Autom. And Intell. Manufact. Conf.: 789-800.

[13] T. Lu, Y. Yih, An Agent-Based Production Control Framework for Multiple-Line Collaborative Manufacturing, 2001, International Journal of Production Research, 39/10: 2155-2176.

[14] Y. Lee, S. Kumara, K. Chatterjee, Multi agent based Dynamic Resource Scheduling for Distributed Multiple Projects Using a Market Mechanism, 2003, J. of Intell. Manufact., 14/5: 471-484.

[15] A. Márkus, T. Kis, J. Váncza, L. Monostori, A Market Approach to Holonic Manufacturing, 1996, Annals of the CIRP, 45/1: 433-436.

[16] J. Váncza, A. Márkus, An Agent Model forIncentive-based Production Scheduling, 2000, Computers in Industry, 43/2: 173–187.

[17] P. Valckenaers, H. Van Brussel, Holonic Manufacturing Execution Systems, 2005, Annals of the CIRP, 54/1: 427-432.

[18] Van Brussel, H., Wyns, J., Valckenaers, P., Bongaerts, L., Peeters, P., Reference architecture for holonic manufacturing systems: PROSA. 1998, Computers In Industry, 37:255-274.

[19] Anussornnitisarn, Pornthep , Nof, Y. Shimon, Etzion, Opher , Decentralized control of cooperative and autonomous agents for solving the distributed resource allocation problem, 2005, Int. J. Production Economics 98, 114–128.

[20] D. Xue, J. Sun, D- H. Norrie, An Intelligent optimal production scheduling using constraint-based search and agent-based collaboration, 2001, Computers in Industry 46, 209-231.

[21] Sergio Cavalieri, Marco Garetti, Marco Macchi, Marco Taisch, An experimental benchmarking of two multi-agent architectures for production scheduling and control, 2000, Computers in Industry 43, 139-152.

[22] Tseng, M. Mitchell, Ming Lei, Chuanjun Su, A Collaborative Control System for Mass Customization Manufacturing, 1997.

[23] A.I. Anosike , D.Z. Zhang , An agent-based approach for integrating manufacturing operations, 2006, Int. J. Production Economics.

[24] J. Sun, D. Xue, Dynamic Reactive Scheduling for responding to changes of production orders and manufacturing resources, 2001, Computers in Industry 46 189-207.

[25] Y. Dumond , C. Roche, Formal specification of a multi-agent system architecture for manufacture: the contribution of the p-calculus, 2000,Journal of Materials Processing Technology 107,209-215

[26] Shehory, Onn , Kraus, Sarit, Yadgar, Osher, Emergent cooperative goal-satisfaction in large-scale automated-agent systems, 1999, Artificial Intelligence, 110, 1–55.

[27] Shen, Weiming , Hao. Qi, Hyun Yoon, Joong, Norrie, Douglas H., Applications of agent-based systems in intelligent manufacturing: An updated review, 2006, Advanced Engineering Informatics 20 415–431.

[28] Ana Madureira, Joaquim Santos, Nuno Gomes Fernandes, Carlos Ramos, 2007, Proposal of a Cooperation Mechanism for Team-Work Based Multi-Agent System in Dynamic Scheduling through Meta-Heuristics, 2007IEEE International Symposium on Assembly and Manufacturing (ISAM07), Ann Arbor (USA), pp. 233-238, ISBN: 1-4244-0563

3D Simulation of Suicide Bombing – Using Computers to Save Lives

Zeeshan-ul-hassan Usmani
Department of Computer Science
Florida Institute of Technology
zusmani@fit.edu

Eyosias Yoseph Imana
Department of Electrical Eng.
Florida Institute of Technology
eyoseph2007@fit.edu

Daniel Kirk
Department of Aerospace and Mech. Eng.
Florida Institute of Technology
dkirk@fit.edu

Abstract – ***This paper asks a new question: how, as computer scientists, we can save the lives of the people living in the troubled areas of the world where suicide bombing and improvised explosive devices (IEDs) explosions become routine incidents in daily lives. How we can utilize the power of modeling and simulation and the advantages of agents based simulation to create a virtual environment to run and test the suicide bombing incidents. And to gain valuable insights on geometrical arrangements, and crowd formations and densities to find out the optimum and safest crowd formations. Which when followed will minimize the number of deaths and injuries during a suicide bombing attack. This paper presents the science of suicide bombing under the framework of agent based simulation. It also explains the physics, explosive models, mathematics and the assumptions we need to create such a simulation. The work also describes human shields available in the crowd with partial and full coverage in both 2 dimensional and 3 dimensional environments. A virtual simulation tool has been developed which is capable of assessing the impact of crowd formation patterns and their densities on the magnitude of injury and number of casualties during a suicide bombing attack. Results indicated that the worst crowd formation is street (Zig-Zag) where 30% crowd can be dead and 45% can be injured, given typical explosive carrying capacity of a single suicide bomber. Row wise crowd formations was found to be the best for reducing the effectiveness of an attack with 18% crowd in lethal zone and 38% in injury zones. For a typical suicide bombing attack, we can reduce the number of fatalities by 12%, and the number of injuries by 7% by simply following the recommendations in this paper. Simulation results were compared and validated by the real-life incidents and found to be in good agreement. Line-of-sight with the attacker, rushing towards the exit, and stampede were found to be the most lethal choices both during and after the attack. These findings, although preliminary, may have implications for emergency response and counter terrorism.***

I. INTRODUCTION

Suicide bombing is an operational method in which the very act of the attack is dependent upon the death of the perpetrator [12]. A suicide attack can be defined as a politically motivated and a violent-intended action, with prior intent, by one or more individuals who choose to take their own life while causing maximum damage to the chosen target. Suicide bombing has become one of the most lethal, unforeseeable and favorite modus operandi of terrorist organizations. Though only 3% of all terrorist attacks around the world can be classified as suicide bombing attacks, these account for 48% of the casualties [12].

Past research has focused on developing psychological profiles of suicide bombers, understanding the economical logic behind the attacks [6,7,8], explaining the strategic and political gains of these attacks, their role in destabilizing countries [2, 5], and the role of bystanders in reducing the casualties of suicide bombing attacks [8, 10]. The specifics of the actual crowd formation and orientation of the bomber with respect to the crowd has not been examined. The presented simulation examines variables such as the number and arrangement of people within a crowd for typical layouts, the number of suicide bombers, and the nature of the explosion including equivalent weight of TNT and the duration of the resulting blast wave pulse for both 2D and 3D environments. The goals of the analysis are to determine optimal crowd formations to reduce the deaths and/or injuries of individuals in the crowd, to determine what architectural and geometric changes can reduce the number of casualties and injuries, and what is the correlation between variant crowd densities and formations with the weight and pulse duration of the explosives? The main objective of our research is to explore and identify crowd formation precautions that when followed will minimize the number of deaths and injuries during a suicide bombing attack.

II. EXPLOSIVE MODEL

In order to model the effects of a suicide bomber on a given crowd formation, it is essential to properly model the deleterious properties of the blast waves themselves. A conventional bomb generates a blast wave that spreads out spherically from the origin of the explosion. The strength of the blast wave decreases exponentially with distance [13]. Although the physics of blast waves are complex and nonlinear, a wave may be broadly characterized by its peak overpressure (pressure above atmospheric) and the duration of the positive phase of the blast event.

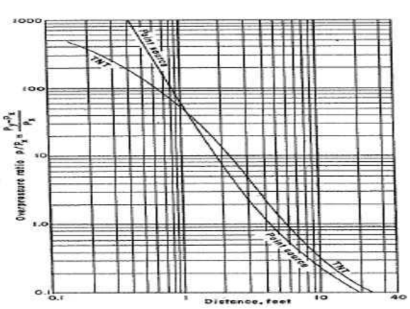


Figure 1: Ideal Overpressure Vs Distance Curve [9].

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_88, © Springer Science+Business Media B.V. 2010

Experimental and theoretical means have been used to obtain important parameters associated with blast waves. A theoretical analysis for peak overpressure utilizes the same mathematical approach as for a planar shock wave, but includes the effects of spherical divergence and the transient nature of the blast event [3, 9]. As an example, values for the peak overpressure generated in a standard atmosphere for the blast wave generated from a one pound spherical charge of TNT are shown in Figure 1. Also shown is the peak overpressure that would be expected at various distances had the energy released by the one pound point source of TNT been concentrated into a point source.

In order to apply the behavior depicted in Figure 1 for any weight of TNT, scaling laws for explosions based on geometrical similarity are used. Two explosions can be expected to give identical blast wave peak overpressures at distances which are proportional to the cube root of the respective energy release [9]:

$$\frac{d}{d_0} = \left(\frac{W}{W_0}\right)^{\frac{1}{3}} \tag{1}$$

The energy release factor is contained in a ratio $(W/W_0)^{1/3}$, where W is the energy release, or amount of TNT, in the explosion to be described, and W_0 is that of a reference amount of TNT, such as the one pound explosion shown in Figure 1. By using this scaling law, the distance at which a given peak overpressure is produced by a reference explosion may be scaled up or down to provide a corresponding distance for other explosions.

Impulse is also an important aspect of the damage-causing ability of the blast, and may become a controlling factor for short duration, small yield explosives. The significant portion of the impulse is associated with the positive phase. The decay of blast overpressure does not follow a typical logarithmic decay relation, because the overpressure drops to zero in finite time (Figure 1). A quasi-exponential form, in terms of a decay parameter, α, and of a time, t, which is measured from the instant the shock front arrives, the pressure can be given as:

$$p = p_0\left(1 - \frac{t}{t_d}\right)e^{\frac{-\alpha t}{t_d}} \tag{2}$$

Where p is the instantaneous overpressure at time t, p_0 the maximum or peak overpressure observed when t is zero, and, t_d, the time duration. The decay parameter is also a measure of intensity of the shock system. Equation (2) may also be used in the simulation if the decay parameter, α, is specified, for example to determine the evolution of the positive phase duration as a function of distance from the explosive center.

In order to tie together the influence of peak overpressure and duration to injury and fatality probability, a series of data curves were utilized. Figure 2 shows the fatality curves predicted for 70-kg man applicable to free-steam situations where the long axis of the body is perpendicular to the direction of propagation of the shocked blast wave.

The inputs to the simulation for the suicide bomber are the equivalent amount of TNT that the bomber is carrying and the initial duration (or α if equation (2) is used) of the blast wave at the bomber's location. Specifying the amount of TNT, using the scaling law of equation (1), and the overpressure versus distance curve of Figure 1, then allows for the calculation of the peak overpressure at any distance away from the bomber. Using this peak overpressure and the increasing duration given by the scaled baseline data set a new duration of the blast wave can be calculated at any distance away from the bomber. Using these two pieces of information and injury or fatality probability curves, such as Figure 2, an estimate of the injury or fatality levels at any location of the bomber can be calculated for various crowd formations.

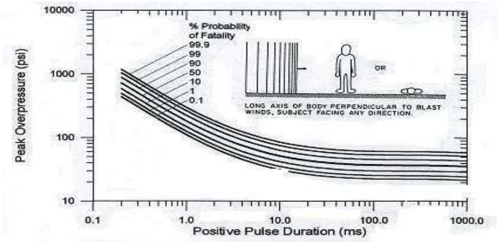


Figure 2: Fatality curves as a function of blast wave peak overpressure and positive pulse duration [3].

III. BLOCKERS AND ZONES

Blockage or human shields present in the crowd can play an important role in the event of suicide bombing attack. A person in front of you, and thus providing a blockage in the line-of-sight between you and the suicide bomber, can actually save your life by taking most of the shrapnel coming from the explosion or by consuming the part of the blast wave overpressure PSI. Spatial distribution of individuals in the crowd can significantly alter the casualty toll, and different crowd formations can yield different outcome with the same amount and kind of explosives.

Persons on the line of sight between a given target and the blast point are termed as full blockers. The partial blockers are those who are not on the line of sight but their body width can cover some part of the body of the person from the blast projectiles. To the best of our knowledge, this study is the first instance of introducing partial blockers in the blast wave simulation. Figure 3 presents the blockage model. Each person in the area is modeled by a line segment, where mid-point is the position of the person and the length is the specified width of the person. Each line in the model is represented by the coordinates of the two end points. The line between the mid-point of the target and the blast point is called the line-of-sight. Each target is represented by a line segment termed as body-width-line. The triangle whose base is the body-width-line of the target and opposite vertex is the blast point and is termed as blast triangle.

The line segment between the blast point (b_1, b_2) and the center of the target (t_{01}, t_{02}) is constructed and its slope is calculated. Assuming all the people are going to face towards the blast point at the moment of blast, the body-width-line of the target will be perpendicular to the line of sight. The slope of this line is minus of the slope of the line of sight. Using simple coordinate geometry, one can easily determine the end points of the body-width-line of the target $((t_{11},t_{12}):(t_{21},t_{22}))$ having the mid-point of the line ((t_{01},t_{02})), the body width and the slope of the line. Finding the end points of the body-width-line of the target, we can easily construct the two other sides of the blast triangle. All other people's body-width-line is assumed to have the same slope as the slope of the body-width- line of the target. Taking this slope and the position coordinate and the width, it is trivial to determine the end points of the body-width-line of each person.

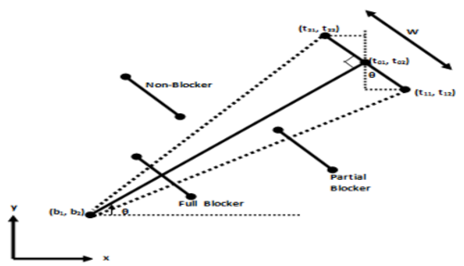


Figure 3. Full, Partial and No Blockers

It is also worth noting that all infinite slopes are approximated by $\pm 1\times 10^6$ in the code. To determine the blockage we have to see if the body-width-line (representing a person) is intersecting with either the line-of-sight or the sides of the blast triangle. If a body-width-line is intersecting the line of sight, the person represented by this line is taken as full blocker. Else, if it is intersecting with either of the sides of the blast triangle, the person will be considered as a partial blocker. Otherwise the person has no interaction with the target.

To find blockers in 3 dimensional, a Cartesian(x-y-z) plane is used as a reference to the distribution of agents. Each agent is modeled by a four sided polygon whose dimensions are determined by the height and width of the agent. These polygons are made to lie parallel to the y-z plane to reduce the computational overhead.

Figure 4 illustrates the concept. There are four planes which enclose the cone whose vertex is the bomber position and whose base is the 4 sided polygons modeling the agents. The cone is called the blast cone and the enclosing planes are termed as blast cone planes. The plane containing this polygon is called the agent body plane and the polygon is called the agent body polygon. The four lines segments extending from the bomb position and the corner points of the polygon are called the blast lines. The algorithm consecutively considers all of the agents as a target and checks if the other agent is interfering with this target agent from the blast point or not. It determines the blockage of an agent if the other agent polygon is intruding into the blast cone. A blocker is called a full blocker if it is intersecting the line of sight between the bomber and the target agent and called a partial blocker otherwise.

To check if an agent is intruding into the blast cone, first we find the smallest distance between the line of sight and the blast lines from the position point of the agent and the bomb. If this distance is less than half of the width of the agent, the line apparently crosses the body plane between the polygon sides and the agent will be considered as a blocker. If the line is the line of sight, it will be the full blocker and if the line is one of the blast lines, it will be considered as a partial blocker. If the smallest distance from each of the lines obtained is greater than half of the width the agent it is not a blocker.

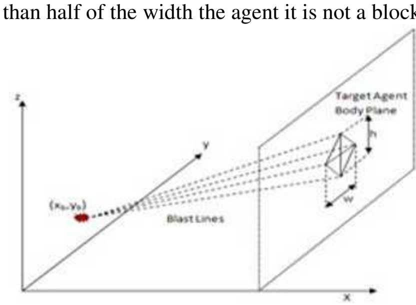


Figure 4. Finding Blockers in 3D

Injuries that occur as a result of explosions can be grouped into several broad categories. We have divided our results in six zones, three for lethality and three for injuries. Lethal zone one refers to 99% probability of death, lethal zone two represent 50% probability of death and the zone three counts the persons with 1% probability of death. Similarly, injuries are divided into three zones. Injury zone one represents the persons who are getting 60 or more overpressure PSI, zone two refers to more than 40 and less than 60 overpressure PSI and zone three for more than 20 and less than 40 overpressure PSI. In general, 60 PSI means severe injury like missing body parts, amputation, brain or heart rupture or Abbreviated Injury Score (AIS) 3. 40 PSI usually refers to the rupture of air-filled organs like lungs and kidney or AIS 2, and 20 PSI is usually responsible for minor bruises and ear-drum rupture or AIS 1. Persons below the range of 20 PSI are generally unharmed.

Lethal Zones	No Blocker	Full Blocker	Partial Blocker
99%	Dead 99%	Dead 50%	Dead 99%
50%	Dead 50%	Unharmed	Dead 1%
1%	Dead 1%	Unharmed	Unharmed
Injury Zones			
60 PSI	Injured 60 PSI	Injured 20 PSI	Injured 40 PSI
40 PSI	Injured 40 PSI	Unharmed	Injured 20 PSI
20 PSI	Injured 20 PSI	Unharmed	Unharmed

Figure 6. Full and Partial Blockers Impact

Self-explanatory Figure 6 provides the details of the respective impacts of the full and partial blockers on the lethal and injury zones. For example, a person within the 50% lethality zone blocked by a full blocker will be unharmed or the same person blocked by a partial blocker will be downgraded to the lethal zone 3 (1% probability of death).

IV. SIMULATION

The simulation is programmed in Visual C++. We choose the Visual C++ programming language due to its extensive library of graphics and geometry functions (to generate the Cartesian grid with agents) and exceptional coverage of code integration with other third party tools like MatLab® (to code the blast overpressure and explosion models). In the simulation, it is assumed that the crowd is uniformly distributed throughout the area. The explosive range is determined by its weight. Specific simulation inputs are the number of individuals in the vicinity, walking speed of the attacker, time associated with the trigger, crowd formation, pulse duration and the total weight of TNT that is detonated. Additionally the arrival time of the explosive pressure front to travel from the bomber to any given location may also be calculated within the simulation.

We have considered mostly "open space" scenarios to serve as the basis for our crowd formation types (e.g., mosques, streets, concerts etc.). Type of injury caused by overpressure depends on whether overpressure occurs outdoor in open air or within buildings and whether they cause collapse of a building or other structure.

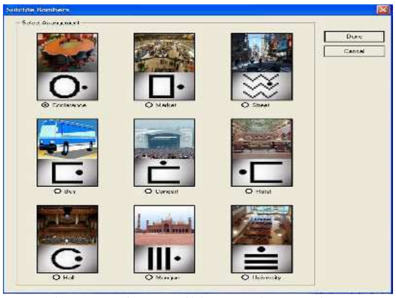


Figure 7: Nine Possible Crowd Formations

There are nine different settings a user can choose from the simulation main screen to estimate the outcome of an attack for a particular crowd formation. There are formations for Conference, Market, Street, Bus, Concert, Hotel, Shopping Mall, Mosque and University Campus. These nine settings were derived from the findings of Mark Harrison, where the majority of the suicide bombing attacks from November 2000 to November 2003 in Israel, occurred on the Streets, Cafeterias, Buses or other open spaces [8]. Users can also define number of participants (victims), number of attackers (suicide bombers), bomb strength (TNT weight in pounds), and bomb-timer (if any). Figure 7 shows the selection menu for crowd formation styles, and Figure 8 shows the display after the blast is simulated.

The display depicts the casualties by red colored icons, those with injuries in light red colored icons, and those who remain unharmed in the attack are shown in blue colored icons. Thus, there are three states of victims after the blast: dead, injured and unharmed (but in panic and contributing in stampede).

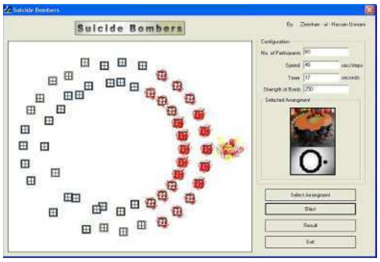


Figure 8: Simulation Screen After the Blast

There are three models of the simulation: Level 0, Level 1 and Level 2. Level 0 (L0) is the basic simulation of blast wave without blockage (full or partial), Level 1 (L1) model is with the full and partial blockage in 2 dimensions and the Level 3 (L3) is the full simulation with partial and full blockage in 3 dimensions (incorporating the height and width of the agents). We have divided the models to examine the exact impact of blockers and switching from 2D to 3D environment in the end results.

V. RESULTS AND VALIDATION

We run the simulation for average case loss scenario for all of the models (L0, L1 and L2). The weight of the explosives being used in the simulation for following results ranges from 1 lb to 30 lbs. The numbers of participants were from 20 to 100 and the pulse duration used for each set of simulation runs were from 0.5 milliseconds to 2 milliseconds. We have also run the simulation for sensitivity analysis for bigger crowds like with 500 to 1000 participants. The overall impact of blast on number of participants get stabilized as the participants increases. For example, the average of participants in the lethal zone were 10 with 20 total participants (50%) and 358 for total participants of 1000 (35.8%). These findings are parallel with Moshe Kress findings in [10]. The simulation was performed for all nine crowd formations with same number of participants and same weight of explosives. The height, weight and number of participants were exactly the same for each run for all three models of the simulation. Following are the average results of 200 simulation runs for each crowd formation with different explosive mass, pulse duration and the number of participants.

We were expecting the upper bound of the results with L0 model, since there is no blockage shield available to people in the crowd, so the model should report more injuries and deaths. For L1 we were expecting the lower bound of the results, since in two dimensions anyone in the line-of-sight can provide blockage, thus minimizing the impact of blast wave overpressure to the people behind the shields.

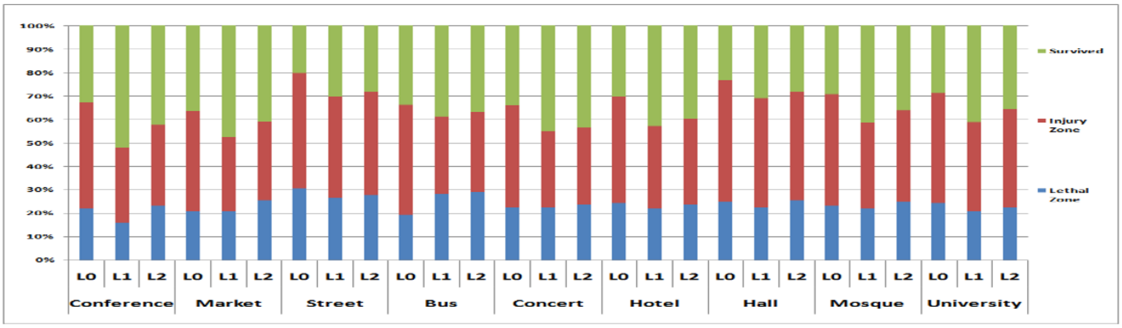


Figure 9. Casualties and Crowd Formations

Our expectations for L2 model results were in between L0 and L1, it should be lower than L0 since it is providing blockage shields to the crowd and it should be greater than L1 due to its three dimensional capabilities. For example, a child standing in front of an adult person in 2D simulation can provide the full blockage while he will be providing only partial blockage in 3D simulation model.

Figure 9 summarizes the findings of the percentages of the people in the lethal and injury zones with given crowd formations. Each set of three bars in Figure 9 represents a crowd formation. It is clear to see that L1 model with blockers have less number of dead and injured people than L0 (without blockers) and L2 has the higher number of death and injuries as compared to L1. The behavior of L2 results makes it more realistic and close to real life situations. We have also performed the simulation runs with 40 and 50 lbs of explosives (though it is uncommon to see the pedestrian suicide bombing attack of that magnitude). The relationship between the increase in the percentage of casualties and injuries with the amount of explosive is observed to be piecewise linear. This relationship is logical since augmenting the explosive material will increase the overpressure pounds per square inch (psi) in the vicinity.

The average deadliest crowd formation for casualties is found to be the Street (Zig-Zag) scenario, where 30% of the participants were in lethal zone and 45% in injury zones. Row wise crowd formations like Market were found to be the best for reducing the effectiveness of an attack, with on average 18% crowd in lethal zone and 38% in injury zones. Thus by only changing the way crowd forms, we can reduce the deaths by 12% and injuries by 7% on average. This is really useful where we have the control to form the crowd, like on airports by placing them in queues, banks, cafeteria, and stadium or in the presidential debates or political rallies. One of the reasons of that dramatic change in the casualties is that in row wise formations, there are fewer people in the direct line-of-sight with the bomber and more people also provide the human shield to others by blocking the blast wave.

To validate our results and to see how close they are with the real-life incidents, we have compiled the database of every single suicide bombing attack in Pakistan from January 1995 to October 2008, 120 incidents in total. Figure 10 shows the comparison of the average number of persons killed and injured in all of the simulation runs against the suicide bombing attacks in Pakistan.

Clearly, the Level 2 3D model with blockers is more close to real-life results than Level 0 with no blockers and Level 1 of blockers in 2D. The average injury per fatality ratio in real-life incidents is 2.18, that is, for every dead person there are 2.18 injured people. The number is pretty much consistent in the history of modern world, where we have 2.6 injuries per fatality ratio in Vietnam War, 2.8 in Korean War, 1.8 in World War I, and 1.6 in World War II. Simulation models on the other hand had produced 1.9 Injuries per fatality in L0 model, 1.6 for L1 model and 1.54 for L2 model. And there is couple of reasons for that. First the current simulation does not count secondary and tertiary blast injuries by fire, debris, fragmentation and shrapnel. Second, the current simulation only accounts for TNT explosive, while in the real-life there are quite a few mixtures of explosives being used. For example, RDX and TNT mixture in the recent suicide bombing attack in Pakistan that claims the life of former Prime Minister Benazir Bhutto and the mixture of Ammonium Nitrate and RDX in Oklahoma City bombings. Third, simulation is not giving the exact number of dead and injured people; instead it is giving the number of people in the lethal and injury zones based on their probabilities of death and injury. For example, a person in the lethal zone 3 with 1% chances of being dead is most likely to be injured and not dead, similarly a person in Injury zone 3 with 20 PSI can be unharmed. There are demographical, environmental and physical characteristics as well, that play an important role in the overall toll. For example, an infant next to the fire cracker can die while a muscular six and half feet person with 250 lbs of weight can survive 1 pound of TNT explosion.

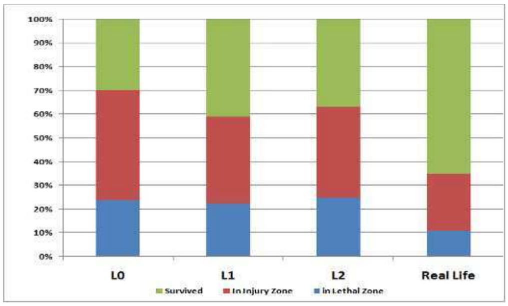


Figure 10. Models Comparison

We expect more realistic results with the incorporation of non-human shields, refraction waves, secondary and tertiary blast injuries and physical characteristics. However, this simulation can provide a good upper bound, lower bound and medium estimates of the number of dead and injured for emergency preparedness, triage of patients and the required number of medical and ambulances facilities for such an event.

The results are in good agreement for the death count but little bit off for injury counts. Beside the aforementioned reasons, one of the reasons for this difference can be totally political, where governments are tend to show the manipualted figures to minimize the aftereffects (like riots, revenge etc) by victim's supporters or a huge cryout in the home state. For example, 4,000 soldiers have been died in Iraq so far since the invasion of the country by US forces in October 2003. Media only concentrate on the dead, while, little has known about more than 250,000 injured soldiers [14]. An injured soldier costs atlesat three times more than the dead soldier economically to the country. Government has to pay disability and social security allownces, and it is a loss of one worker from the labor force, thus a loss of one statical value of life, and the injured also need a caretaker, therefore another loss of the statistical value of life. Given the current geo-political conditions of the world and the US on going war in Iraq and Agahnistan, it is more necessary than ever to examine and employed the technologies to reduce the rate of injured and the dead. Another reason of the gap in the number of injured might be the level of injuries – a victim who has the minor injury and was able to walk may not have been included in the actual count of the injuires in the reallife events.

We have also performed the sensitivity analysis for all of our models. L0 or the basic model results are the same as L1 2D model without blockage. And the L1 2D model without blockage results are similar to the results of L2 3D model without blockage. We recommend to use L0 basic model if there is no need of considering the blockage. L0 can also gives the upperbound of body count. If blockage has to be considered, we recommend using L1 2D model, since L2 3D model contribution is statistically insignificant if only considering the blockage in the crowd, while, 3D demands more computational power and resources. We suggest to use L2 3D model when there is a need of blockage with uneven surfaces like stage or stadium, and the user has to work on bomb fragments, shrapnel, projectiles, and secondary and tertiary blast injuries. 3D model is more realistic when used with the majority of blast characteristics. For the simple estimates L1 2D model is as good as 3D, while, the L0 bsaic model can be used for quick estimation for the required number of medical and emergency management facilities.

Announcing the threat of suicide bombing in the crowd can only make the condition and the causality toll much worse. People will panic and thus increase the possibility of more victims in the line-of-sight with the suicide bomber than before. People will also try to rush towards the exit gates (thus coming closer to suicide bomber in majority of cases) and there will be high chances of a stampede.

VI. CONCLUSIONS AND FUTURE WORK

There are a number of lessons we can learn from the analysis of this suicide bombing simulation with given crowd formation styles. For example, we can reduce the number of fatalities by 12% and number of injuries by 7% by switching the crowd formation from Zig-Zag to Row-Wise formation styles. To avoid the stampede in possible crowd formation, we could arrange more exit points than normally available to such crowds. Suggestions can also be made for architectural design changes in the buildings to reduce the count, for example by placing entrance and exit gates by *X* feet away from the main venue, victims can be reduced by *Y%* (the values depends on environment, crowd information and the weight of explosive). The results can also help to plan for post-disaster management, for example, how many ambulance and doctors we will need if something like this happen to given crowd or how to direct the crowd to behave or run towards particular exit by announcing it through loudspeakers. In the light of these findings, the crowd can be manipulated in real-life by imposing formation guidelines like queues at the airport or by placing the chairs in particular order that will block the line-of-sight with any perspective attacker with the rest of the crowd.

The simulation and findings are limited in that it only incorporates the primary injuries. Future plans are to add secondary effects (e.g., injuries by fire, debris, etc.) to better approximate the real world environment and provide more valid comparisons with the data of suicide bombing attack aftermaths [11]. We will also add the flexibility to create the user defined crowd formations with variable number of entrances and exits in the future. This paper provides an interesting direction for future research to take in investigating the catastrophic event of the suicide bomber attack in hopes of making the world a safer place.

REFERENCES

1. Air Force, "U.S. Air Force Installation Force Protection Guide", 2004
2. Azam, Jean-Paul, "Suicide-bombing as inter- generational investment", Public Choice. Springer. 122, 177-198, 2005.
3. Cooper, Paul W., "Explosive Engineering", Wiley-VCH, 1996
4. FEMA, "Explosive Blast Manual", Section 4.1. Federal Emergency Management Association Press, 2004.
5. Ganor, Boaz, "The Rationality of the Islamic Radical Suicide Attack Phenomenon", Countering Suicide Terrorism, Institute of Counter Terrorism, 2000.
6. Gupta, Dipak K. & Mundra, Kusum, "Suicide Bombing as a Strategic Weapon: An Empirical Investigation of Hamas and Islamic Jihad. Terrorism and Political Violence", 17, 573-598, 2005.
7. Harrison, Mark, "The Immediate Effects of Suicide Attacks", University of Warwick, 2004.
8. Harrison, Mark, "Bombers and Bystanders in Suicide Attacks in Isreal 2000 to 2003", Studies in Conflict and Terrorism. 29, 187-206, 2006.
9. Kinney, Gilbert & Graham, Kenneth, "Explosive Shocks in Air", 2nd Sub edition. Springer, 1985.
10. Kress, Moshe, "The Effect of Crowd Density on the Expected Number of Casualties in a Suicide Attack", Wiley Periodicals, 2004.
11. Lester, David, Yang, Bijou & Lindsay, Mark. "Suicide Bombers: Are Psychological Profiles Possible?", Studies in Conflict and Terrorism. 27, 283-295, 2004.
12. Pape, Robert A., "Dying to Win: The Strategic Logic of Suicide Terrorism", Random House, 2005
13. Irwin, R. J., Lerner, M. R., Bealer, J. F., Mantor, P. C., Brackett, D. J., and Tuggle, D. W., "Shock after blast wave injury is caused by a vagally mediated reflex", Journal of Trauma, Volume 47, p. 105-110, 1999.

Stiglitz, Joseph and Bilmes, Linda, "Three Trillion Dollar War, Norton and Company, July 2008

Reliability Investigation of a Hybrid Fuel Cell Electric Vehicle Powered by Downsized Fuel Cells

Pardis Khayyer, Afshin Izadian and Parviz Famouri
West Virginia University
Morgantown, WV 26506 USA

***Abstract*- Fuel cell electric vehicles are mostly relying on operation of their fuel cell and battery system. Single power source systems use battery units as backup; however, in heavy loads or instances with low State Of Charge (SOC) levels, there is a need for other mechanisms to provide reliable energy for the system. This paper investigates the natural enhanced reliability of operation in an advanced system configuration with two downsized fuel cells. The two fuel cell configuration brings high fuel efficiency by economic load sharing between two fuel cells. The reliability of this system configuration is investigated and compared with conventional designs of hybrid fuel cell vehicles.**

Keywords: Hybrid Fuel Cell Vehicle, Reliability, Double Fuel Cell.

I. INTRODUCTION

Recent advancements in fuel cell technology have made them suitable candidates for portable applications and specifically in vehicles. Electric vehicles take the most advantage of fuel cells in different configurations. They bring the advantage of longer driving range compare to battery-powered vehicles and increase the performance of the system by providing fuel-economy configurations. Electric vehicles use fuel cells either as the only source of energy or they are connected to a backup source of power such as battery or ultracapacitors to configure non-hybrid or hybrid fuel cell electric vehicles respectively. Traditional designs of hybrid electric cars utilize a single fuel cell and battery backup. The new configuration of two fuel cell power sources in hybrid fuel cell vehicles was proposed in [1]. Fuel cells were downsized to provide the same amount of power and offered the advantage of a highly fuel economic design. Highly efficient driving conditions in urban applications were obtained. Figure 1 shows the double downsized fuel cell vehicle topology.

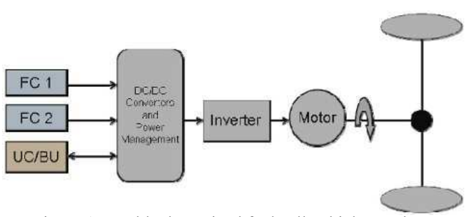


Figure 1, Double downsized fuel cell vehicle topology

One of the major obstacles in commercialization of fuel cell vehicles is their low reliability of operation [2] which results in a huge reduction in the overall system reliabilities. [2] and [3] have conducted research on overall reliability of fuel cell vehicles. Reliability analysis of a single PEM (Proton Exchange Membrane) fuel cell system is studied in [4] and [5]. [6] has focused on the reliability analysis of a single Direct Methanol Fuel Cell.

Reliability of a single source fuel cell system highly depends on the operation of the source. Reliability of a fuel cell power source follows the Weibull distribution that decreases over time. This means that an aged fuel system is less reliable. In addition, backup sources such as battery units in electric vehicles are designed for transient operations and cannot provide energy for long time to act as a reliable source. This paper investigates the reliability of operation of a two fuel cell source without considering the contribution of the battery unit since it is designed for a short period of operation.

In the following sections, reliability enhancement of a double fuel cell power source is compared with the reliability of traditional designs. The reliability is introduced and used for various system configurations.

II. RELIABILITY

Reliability by definition is "the probability that an item will perform its function adequately for the desired period of time when operated according to specified conditions" [7]. Reliability is defined mathematically as

$$R(t) = \int_{t}^{\infty} f(x)dx, \tag{1}$$

where $f(x)$ is the failure probability density function and t is the time period. $F(t)$, the cumulative probability distribution function which is also called the failure probability, is defined as

$$F(t) = \int_{-\infty}^{t} f(x)dx, \tag{2}$$

In reliability analysis, it is assumed that the system have the chance to operate without failure during a specific period of time. A fuel cell is a two state device which either operates or fails during its operation. Configuration of components in the system can categorize a network into four distinctive cases of series, parallel, k-out-of-m unit network and standby redundant system. These systems are defined for the convenience of readers.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_89,

A. Series Network

If the devices/subsystems of a network are connected in series, it is called a series network. In such network, failure of any of the components leads to failure of the whole system. In the series network, overall reliability of this network consisting of k components is calculated by

$$R_s(t) = \{1 - F_1(t)\} \times \{1 - F_2(t)\} \times \{1 - F_3(t)\} ... \{1 - F_k(t)\}, \quad (3)$$

where $R_i(t)$ is the i^{th} unit/component reliability and $F_i(t)$ is the i^{th} component failure probability for $i = 1,2,....k$ and is defined as,

$$\{1 - F_i(t)\} \equiv R_i(t). \quad (4)$$

B. Parallel Network

If the devices/subsystems of a network are connected in parallel, the network is called a parallel network. This type of system only fails when all its components/subsystems fail to operate; therefore, this configuration is used to increase reliability of the overall system. The reliability of the overall system is calculated by

$$R_p(t) = 1 - \prod_{i=1}^{k} F_i(t). \quad (5)$$

C. k-out-of-m Unit Network

This type of network contains *m* units and operates if *k* parallel units operate. The reliability is defined as

$$R_{k/m}(t) = \sum_{i=k}^{m} \binom{m}{i} [R(t)]^i [1 - R(t)]^{k-i}, \quad (6)$$

where $R(t)$ is the unit reliability, *m* is the total number of system units and *k* is the number of units required for the function of the system.

D. Standby Redundant System

In a standby redundant system, *k* units are on standby while one unit functions. The system reliability in this case is calculated by

$$R_s(t) = \sum_{i=0}^{k} \left\{ \int_0^t \lambda(t).dt \right\}^i e^{-\int_0^t \lambda(t)dt} (i!)^{-1}, \quad (7)$$

where λ is the hazard rate or instantaneous failure rate and is defined as "the rate of change of the failed components quantity divided by number of survived components at time *t*" [7].

III. Weibull Distribution

Different distributions such as Weibull, Normal, Exponential Uniform, Extreme value, etc. have applications in reliability engineering analysis. Weibull distribution is commonly used for reliability analysis related to PEM fuel cell systems [5]. The hazard rate for the Weibull distribution is defined as,

$$\lambda(t) = At^B. \quad (8)$$

More details are defined as

$$\lambda(t) = \frac{\beta}{\theta} \left(\frac{t - t_0}{\theta} \right)^{\beta - 1}, \quad (9)$$

where β is defined as the shape parameter, θ is the characteristic life or scale parameter and t_0 is the location parameter. The probability density function of Weibull distribution is expressed as,

$$f(t) = \frac{\beta}{\theta} \left(\frac{t - t_0}{\theta} \right)^{\beta - 1} e^{-\left(\frac{t}{\theta}\right)^{\beta}} \quad (10)$$

The reliability function is expressed as

$$R(t) = e^{-\left(\frac{t - t_0}{\theta}\right)^{\beta}}. \quad (11)$$

Weibull distribution becomes an exponential case when $\beta = 1$ (the failure rate is independent of age). In the next section the system reliability analysis for the proposed configuration of multiple fuel cells in hybrid fuel cell vehicle is presented.

IV. Reliability Analysis for Double Fuel Cell System

In this section, reliability analysis of one fuel cell is formulated and is used for analysis of a parallel network. The probability of a single fuel cell is considered "1" at the start of operation and is decreased as time increases. In (11), β equals 1 [8] which results in a constant failure rate of $\lambda = \frac{1}{\theta}$, and t_0 is zero for a brand new fuel cell unit [9].

Figure 2 shows the reliability of individual fuel cells and the resultant parallel network of double power sources for different failure rate values. This figure illustrates a case where failure rate of fuel cell #1 is more than failure rate of fuel cell 2; therefore, the reliability of fuel cell #1 decays faster. This figure also shows an improvement in the overall reliability of the system which is greater than each of fuel cells individually. This demonstrates higher reliability of the multiple fuel cell configuration proposed in this research.

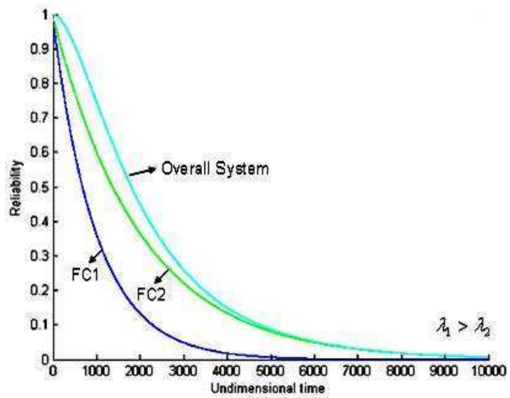


Figure 2, Each fuel cell and the parallel system reliability curves for $\lambda_1 > \lambda_2$

Figure 3 shows the reliability evaluation of fuel cells with the same failure rate values. In this case, both fuel cell reliability curves decay with the same rate; however, the overall reliability of the parallel system is higher than the individual cells.

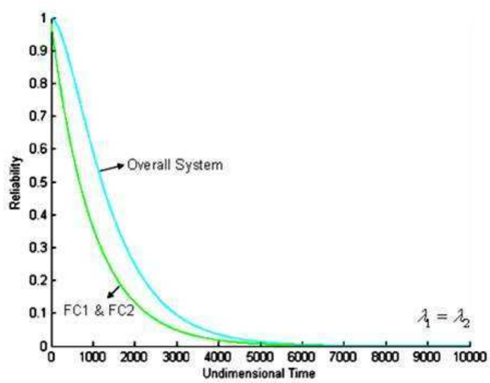


Figure 3, Each fuel cell and the parallel system reliability curves for $\lambda_1 = \lambda_2$

Simulations indicate that by using the multiple fuel cell configuration not only higher efficiency is gained for the system in urban driving cycles, but also a more reliable system is achieved by implementing the new topology.

V. CONCLUSION

In this paper the reliability analysis for the new configuration of a double fuel cell vehicle was discussed. The configuration of double fuel cells in hybrid fuel cell vehicles brings the advantage of high efficiency and fuel economic operation while enhancing the reliability of operation in power sources. Results demonstrated that higher reliability was obtained in the system by downsizing the fuel cell power source.

REFERENCES

[1] P. Khayyer and P. Famouri, “Application of Two Fuel Cells in Hybrid Fuel Cell Vehicles”, *Powertrains, Fuels and Lubricants Meeting, October 2008, Chicago, IL, USA*

[2] M. Marchesoni and S. Savio, “Reliability Analysis of a Fuel Cell Electric Car”. *2005 European Conference on Power Electronics and Applications, 11-14 Sept. 2005. 10 pp.*

[3] L. Wu and H. Li, H, “The Reliability Work in Fuel Cell Vehicle’s Road Test”, *IEEE International Conference on Vehicular Electronics and Safety, 2006. ICVES 2006. 13-15 Dec. 2006. pp. 481 – 484.*

[4] V. Mangoni, M. Pagano and G. Velotto, “Fuel Cell Reliability Model based on Uncertain Data”, *International Conference on Clean Electrical Power, 2007. ICCEP '07. 21-23 May 2007. pp. 730 – 735.*

[5] S. A. Feitelberg,, J. Stathopoulos and Z. Qi, “Reliability of Plug Power GenSysTM Fuel Cell Systems”, *Journal of Power Sources, vol. 147, Issues 1-2, 9 September 2005, pp. 203-207.*

[6] G. Aydinli, N.S. Sisworahardjo and M.S. Alam, “Reliability and Sensitivity Analysis of Low Power Portable Direct Methanol Fuel Cell”, *The International Conference on "Computer as a Tool"EUROCON, 2007. 9-12 Sept. 2007, pp. 1457 – 1462.*

[7] B. S. Dhillon, *“Reliability Engineering in Systems Design and Operation”*, Publisher: Van Nostrand Reinhold Company Inc., 1983.

[8] K. Åström, E. Fontell and S. Virtanen, “Reliability analysis and initial requirements for FC systems and stacks”, *Journal of Power Sources, Volume 171, Issue 1, 19 September 2007, Pages 46-54*

[9] Relex Web Page: *http://www.relex.com/resources/art/art_weibull3.asp*

Dispatching of Active Power Reserve for Higher Reliability of Restructured Power Systems

Mohammad Taghi Ameli[#1], Saeid Moslehpour[#2] ,Ziba Farzadpour[#3],

[#1, #3]Power and Water University of Technology
[#2]University of Hartford, West Hartford, CT, USA

[#1] ameli@pwut.ac.ir
[#2] moslehpou@hartford.edu
[#3] zibafarzad@yahoo.com

Abstract: **In power systems that are restructured on the basis of system reliability or the costs, various methods can be employed for allocating the spinning reserve to various generating units.**
This paper firstly describes two methods for allocation of reserve power:

- **a cost-based method**
- **a reliability-based method**

and secondly, presents a combined method that simultaneously considers both reliability and the costs. After reviewing the results of each method, a new method will be presented that can be applied in the restructured power system. To demonstrate the results of various methods, this research has used the IEEE network model RTS (Reliability Test System) with 24 Bus bars.

Keywords and Phrases: Spinning Reserve (SR), Allocation, Reliability, Cost, Restructuring, Risk

1. Introduction

The main purpose of any power system is to secure the uninterrupted, economical, and reliable provision of electrical energy to satisfy the demands of the small and large consumers (customers). Because of social patterns, life-styles and behavioral patterns, modern societies are dramatically dependent on continuous access to electrical energy.

Because of unpredictable failure or malfunctioning of some system components that are beyond the control of engineers, the continuous and uninterrupted availability of electrical energy requires special equipments that cost money. Increased investment in planning and operating phases may reduce the probability of power outage and interruptions in the supply of energy to consumers.

Excessive capital investment increases the operating costs and necessitates a corresponding boost in the tariffs or schedule of rates; and this violates the economical constraints. Conversely, if investment is below a certain level, then reliability will be decreased that costs the loss of customers. Making a sound decision on reliability and economic constraints is a critical problem for system management in both planning and operating phases and always a main concern of the managers of system operations.

For reliability of power networks in restructured systems, various models are studied, one of which is outsourcing for power system reliability from the market of peripheral services. Important problems that are considered in this market are: reactive power and voltage-regulation services, spinning reserve, non-spinning reserve, frequency control services etc.

In this context and direction, to achieve the desirable levels of the indicators of network reliability, one of the main challenges of the power systems is designing a model to quantify the necessary and sufficient levels of required services and regulate the planning for acquisition of the services.

As an important component of the peripheral services, SR needs special attention, and one of the main problems is the method of determining the optimal level of SR in any power system. Previous research in operating a power system has offered several methods. However, the quantity of SR is not the only factor that influences the system reliability, as the method of allocating the SR to various production units is also important and deserves particular attention.

2. Methods of Allocating the SR

In operating a power system, the operator is strictly required to consider certain constraints. Ignoring these constraints can cause great damage and loss. Reliability is one of the critical constraints that demands particular attention otherwise customer base and profitability will be seriously endangered.

2-1 Risk-Based SR Allocation Method (RBSRAM)

In this method, optimized allocation of SR is achieved by optimizing Unit Commitment Risk(UCR).

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_90,

Using a Capacity Outage Probability Table (COPT), the reserve capacity is calculated on the basis of UCR for various allocations, and the allocation scheme with lowest risk is selected as the allocation that provides highest reliability.

Related constraints are as follows:

(a) The Constraint for Reserve Capacity of the System:

This constraint shows that the sum of spinning reserve of all generating units should be equal to the SR capacity needs of the system.

$$\sum_{b=1}^{N_{bi}} \sum_{g=1}^{N_{gi}} g_{srgbi} = T_{sr} \qquad (2\text{-}1)$$

Where,

N_{gi} is the number of generating companies in the power market for allocation ***i***

N_{bi} is the number of generating buses in allocation ***i***

g_{srgbi} is the available SR from unit ***g*** in generating bus ***b*** in allocation ***i***

T_{sr} is the required capacity of SR for the desired level of system reliability

(b) The Constraint for Reserve Capacity of the Unit:

This constraint shows that SR capacity of each unit should be less than or equal to its maximum output.

$$g_{srgbi} \le g_{srgbi}^{\max} \qquad (2\text{-}2)$$

Where,

$g_{srgbi}^{\max}$ is the maximum SR generated by unit ***g*** in generating bus ***b*** for allocation ***i***

To calculate the risk, we utilize the COPT table. [2].

2-2 Cost-Based SR Allocation Method (CBSRAM)

In restructured environments, cost is the central issue, and the invariant goal of higher profitability persuades all companies to search for methods that minimize the cost. In cost-based allocation method, the objective is selection of the optimal SR allocation from the set of all possible allocations. This is achieved by minimization of the total cost with respect to the limits of SR capacities of various generating units. In this method, risk is not considered. [1]

Once the constraints (2-1) and (2-2) is satisfied, the objective function is formulated on the basis of the cost:

$$\min SRC_i = \sum_{b=1}^{N_{bi}} \sum_{g=1}^{N_{gi}} C_{gbi}\left(g_{srgbi}\right) \qquad (2\text{-}3)$$

Where,

SRC_i Total cost of the SR of the system in allocation ***i***

2-3 Hybrid SR Allocation Method (HSRAM)

In optimal SR allocation, companies and operators should carefully consider reliability otherwise substantial losses, in terms of opportunity costs may incur in the power system.

$C_{gbi}(g_{srgbi})$ cost curve of SR of generating unit g in generating bus b in allocation **i**

But ensuring reliability via observance of its respective constraints is also costly, therefore a trade-off is needed to minimize the total costs, i.e. minimize the sum of SR generation and reliability costs. In other words, the minimum cost of SR generation plus the opportunity costs of the involved risks define the optimal SR allocation. Reliability cost is defined as the lost potential revenue due to outage of generating units.

Provided the constraints (2-1) and (2-2) are observed, the optimization is formulated as follows:[1]

$$\min TC_i = RC(UCR_i) + SRC_i \qquad (2\text{-}4)$$

Where,

TC_i is the total cost

$RC(UCR_i)$ is the reliability cost for SR allocation ***i***

SRC_i is the SR cost for allocation ***i***

The flowcharts of all three methods of spinning reserve allocation are shown in Fig. (1) below.

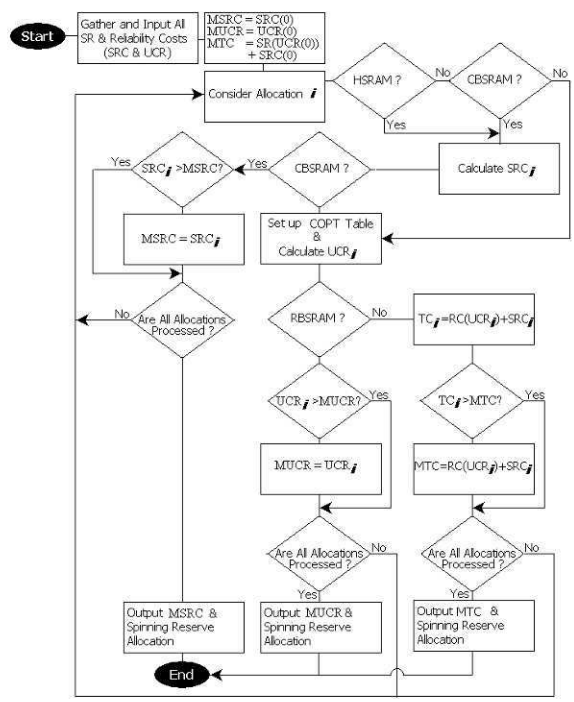


Fig. (1) –Flowchart for Three Algorithms of SR Allocation

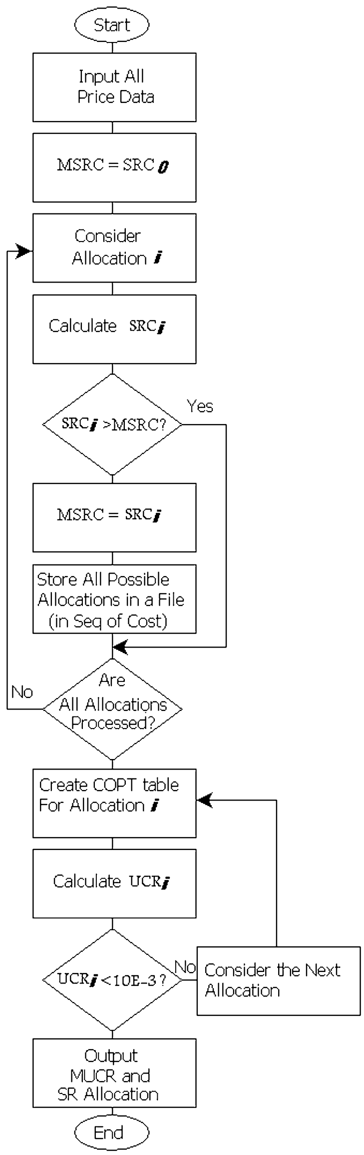


Fig. (2) – Flowchart of SR Allocation in Restructured Systems

2-4 Our Proposed Method for SR Allocation in Restructured Systems

Opportunity cost is not significant in restructured systems, because in such systems, the operator chooses from the set of Market offerings an allocation that has the least cost and at the same time, satisfies the reliability requirements. In this context, a new method for SR allocation in restructured systems is presented in the following. In this method the operator assigns the greatest weight to the cost constraint and evaluates the least cost allocation as the optimal choice. The difference between this method and the cost-based method is that the cost-based method disregards the risks involved in conditions of choosing or selection of the allocation. But in the proposed method, the operator chooses the allocation that firstly, has the least SR generation cost, and secondly, satisfies the reliability constraint.

For this purpose, the operator should consider all offers of SR generation in the market, and prepare a sorted list in ascending sequence of cost. Then he needs to perform the following loop until the criteria for both cost and reliability are satisfied:

Step 1: For the desired level of SR, choose from the least expensive offer(s).

Step 2: Calculate the risk for the chosen candidate(s).

Step 3: If the risks satisfy the reliability criterion, select the chosen candidate(s). Stop■

Step 4: Go back to Step 1 and choose from the remaining offerings.

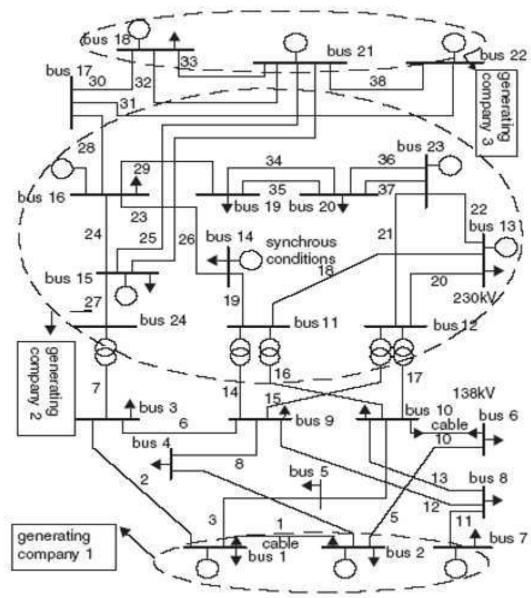


Figure (3) – Schematic Diagram of IEEE-RTS Network

Table (1) – Generating Units in SR Market

Price ($/Mw)	*Capacity (Mw)*	Company	Bus	Unit Number
30	50	3	22	224
30	50	3	22	225
30	50	3	22	226
45	12	2	15	153
45	12	2	15	154
45	12	2	15	155
50	20	1	1	11
50	20	1	2	21
50	20	1	2	22
50	76	1	2	23
50	76	1	2	24
45	155	2	23	231

3. Studies on IEEE Reliability Test System (RTS)

The methods described in this paper were studied using the IEEE RTS test system. [3]

The typical network of IEEE-RTS consists of 24 buses. This network includes 10 generating buses, 10 load buses, 33 transmission lines, 5 transformers, and 32 generating units. Total network capacity is 3405 Mw and the daily peak load is estimated at 2850 Mw.

Table (2) – Various Cases of SR Allocation

Alloc. No.	Company	SR Capacity (Mw)
1	1	20
	2	165
	3	100
2	1	95
	2	190
3	1	135
	3	150
4	1	150
	2	35
	3	100
5	1	215
	2	20
	3	50

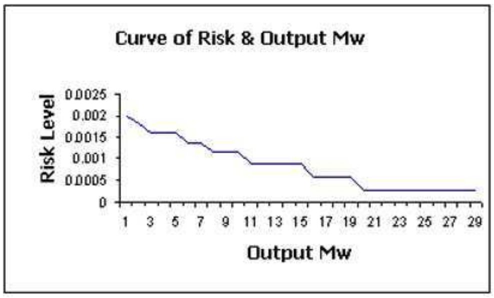


Fig (4) – Risk Curve of the System, for 4th Allocation

The system is divided among 3 generating companies:

1. Generating Company 1 with 11 generators on buses 1, 2, and 7.
2. Generating Company 2 with 13 generators on buses 13, 15, 16, and 23.
3. Generating Company 3 with 8 generators on buses 18, 21, and 22.

Required reserve capacity is estimated at 10% of the peak load, or 285 Mw.

After surveying the energy market 12 generating units were chosen that are shown in Table (1).

All three methods of SR allocation were tested on the IEEE RTS system and the results are as follows:

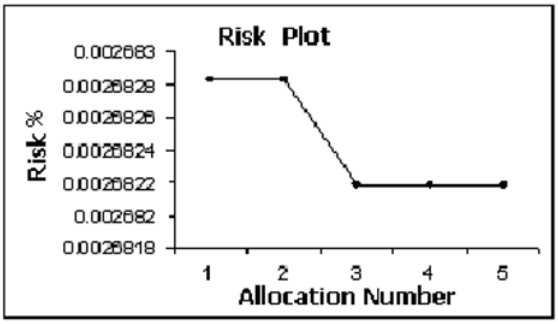
3-1 Risk-Based SR Allocation Method (RBSRAM)

The objective of RBSRAM is to select the allocation with minimum risk from a set of available allocations. The required SR of the system was assumed to be about 10% of the peak load. This method does not require the consideration of cost on selection of the SR allocation. From the 12 remaining generating units in the market, and combining the units, 5 cases of allocation were derived with the following risks (Table 3):

Table (3)- Risk of Various Allocations

Alloc.	Risk%
1	0.002682833769
2	0.002682833814
3	0.002682182671
4	0.002682182488
5	0.002682182791

Fig. (5) – Plot of Allocations' Risk



It can be seen that the 4th allocation has the lowest risk, and according to RBSRAM, this allocation should be selected as the optimum choice. Fig (4) shows the probability that several units go out of service simultaneously is very low.

Table (3) shows the risk associated with various allocations of 285 Mw. The 4th allocation has the lowest risk when the system load is 2850 Mw.

In Fig (5), the allocation numbers and their respective risk are plotted. It can be seen that the allocations 3, 4 and 5 are very close together, while the 4th allocation has the least risk.

3-2 Cost Based Spinning Reserve Allocation Method

For constant SR capacity, if the risk of Unit Commitment is not considered, then the allocation with the least SR cost should be selected.

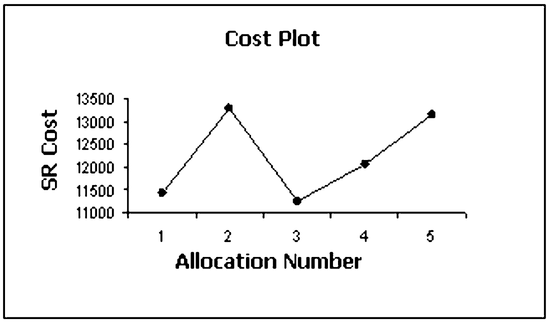
To provide for a constant SR capacity, we can depict 47 different combinations of the generating units. The units in these 47 combinations are grouped in 5 SR allocations.

Using the CBSRAM method, the allocation number 3 that has the least cost is selected. In this allocation, ISO purchases 135 Mw from unit 1, and 150 Mw from unit 3. Table (4) shows the SR cost of various allocations. In Fig (6), the horizontal axis shows the allocation number and the vertical axis shows the cost of providing the SR for each allocation. The cost for allocations 1 and 3 are less than the cost of other allocations, and in allocation 3, the cost is minimum. Though allocation 3 has the minimum cost, but its risk is unacceptable. This risk is equal to the opportunity cost (unrealized profits or the imposed costs. In selection of the optimal allocation, both risk factor and the cost should be minimized, and this is the goal of the Hybrid SR allocation method (HSRAM).

Table (4) – SR Cost for Various Allocations

Allocation No	SR Cost
1	11425
2	13300
3	11250
4	12075
5	13150

Fig (6)- Plot of Allocations' SR Cost



3-3 Hybrid SR Allocation Method (HSRAM)

In this method, the goal of SR allocation is to minimize the sum of reliability cost and the SR cost. It is assumed that the risk cost is modeled by the following linear function.

$$RC(UCR) = 10^7 \times UCR + 5E4$$

Using the HSRAM method, the total cost for each allocation is summarized in Table (5) and is shown in Fig (7). The total cost of allocation number 3 is minimum ($ 88071.82671).

It can be seen that the CBSRAM also produced the same result. This is due to the high reliability of the network that minimizes the risk, and applying the respective risk cost. In general, these two methods may not necessarily produce the same results.

In Fig (7), the horizontal axis shows the allocation number and the vertical axis shows the total costs of SR and risk. The total cost for allocations 1 and 3 is less than the cost of other allocations, and the cost for allocation 3 is the minimum.

Table (5) – Combined costs of SR and Risk

Alloc. Number	Risk	Cost ($)		
		Risk	SR	Total
1	0.002682833769	76828	11425	88253
2	0.002682833814	76828	13300	90128
3	**0.002682182671**	**76821**	**11250**	**88071**
4	0.002682182488	76821	12075	88896
5	0.002682182791	76821	13150	89971

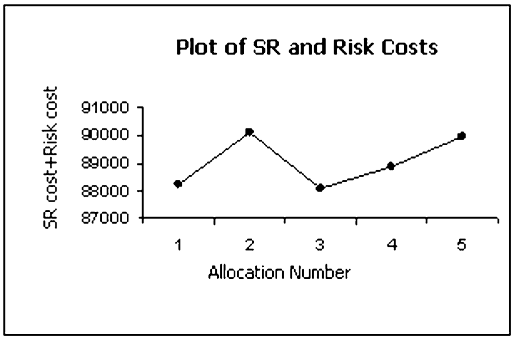


Fig (7) – Plot of SR+Risk costs for Allocations

3-4 The Method of Allocating Spinning Reserve in Restructured Systems

In restructured systems, the goal of SR allocation is to find the allocation with minimum SR allocation cost while its risk is acceptable.

To perform this procedure that is based on algorithm flowchart of Fig (2), the operator firstly sorts in ascending sequence of cost, the list of participating units in the SR market. The proposals of various units for SR costs are shown in the Table (1). These units are sorted in sequence of the costs derived in Table (6).

The least cost combination is when 150 Mw is purchased from the unit 3 and 135 Mw from unit 1. With this selection, the COPT table is prepared for the network under study and the risk is calculated for this case. To ensure that the network reliability constraint is satisfied, a permissible risk level, (α), is defined, and in COPT table, where the available capacity matches the load demand, the value of risk is compared against **α** and the risk should be less than α (risk < **α**). According to references [2] and [4], the permissible value of **α** is 0.001. If the risk is higher than **α**, then the next allocation with a higher SR allocation cost will be considered. Again, the COPT table should be formed on the basis of this allocation and its respective risk is re-calculated…

This sequence of operations are repeated until the condition of (risk < **α**) is satisfied. The first allocation that satisfies this condition will be selected as the optimum choice.

Selection of the 4^{th} allocation yields a risk of 0.002682182488 and all other allocations produce a risk that exceeds the permissible value. Hence, we select the allocation that creates the minimum risk as the optimal allocation thereby reliability of the system is maximized. Using this method, the 4^{th} allocation would be chosen as the allocation that will have the lowest risk.

4. Analysis of the Results

4-1 RBSRAM Method

- Using the RBSRAM method, we obtained an allocation that provides the lowest level of risk. This method can greatly enhance the reliability but in the system under our study, the risk-factors of all cases were within the permissible limit.
- Figure (5) indicates that the allocations 1 and 2 have higher risks, because the units chosen for SR provision had the higher risk of outage.

- Figure (5) also indicates that the system risk for allocations 3, 4, and 5 are close together and allocation 4 has the minimum risk. Since the difference of the risks is negligible, any one of these allocations that has the lowest SR cost can be selected.
- It is seen that the risk of allocations 3, 4, and 5 are lower than that of the allocations 1 or 2. In these three allocations, more units lend themselves to SR provision.
- In allocation number 4 that has the least risk, units with 12 Mw, 50 Mw, and 76 Mw are chosen that have a lower probability of outage.

4-2 CBSRAM Method

- Considering the Fig (6), it can be seen that SR provision cost of allocation 3 is the lowest, and allocation number 1 is next to that.

Table (6) – Generating Units in the SR Market In sequence of Price

Unit No.	Bus	Company	capacity (Mw)	Price ($/Mw)
224	22	3	50	30
225	22	3	50	30
226	22	3	50	30
231	23	2	155	45
153	15	2	12	45
154	15	2	12	45
155	15	2	12	45
11	1	1	20	50
21	2	1	20	50
22	2	1	20	50
23	2	1	76	50
24	2	1	76	50

- According to the results obtained from CBSRAM method, the allocation number 3 was found to have the least cost. In this allocation the units with lowest price were used.

4-3 HSRAM Method

- Fig (7) indicates that allocations 3 and 1 have the lowest combined costs of risk and SR provision.
- It is seen that the cost of SR provision has a higher impact on the final cost, and this is due to high reliability of the system under this study that has reduces the impact of risk in various allocations.

4-4 The method of SR allocation in Restructured systems

- In restructured systems, where the risk or opportunity cost is meaningless and insignificant, a wise operator should search for an allocation that creates the least cost. In this approach, he should also consider the reliability constraint. In other words, he should select an allocation with the least cost that satisfies the permissible risk.

4-5 Comparison of the Results Obtained in Various Methods

- Execution of the RBSRAM algorithm is far more difficult and time consuming than the CBSRAM algorithm. Therefore, in highly reliable networks, like the network in our study, the cost based algorithm CBSRAM is more suitable.
- In systems with low reliability, the RBSRAM algorithm is more suitable as it upgrades the reliability.
- The hybrid algorithm (HSRAM) yields an allocation that has the least total costs of SR provision and risk, and the selected allocation is, in a way, the optimal selection.
- The new method for restructured systems presented in this paper minimizes the costs and satisfies the reliability constraint, and additionally, can be performed very quickly. This method is recommended for cases that the system operator is not concerned about the opportunity costs.
- Due to standard characteristics of the RTS networks, and the low probability of outage that is due to highly reliaility, both algorithms HSRAM and CBSRAM produce identical results.
- The results obtained for restructured systems and those obtained from the cost based method were identical, because all selectable allocations were associated with the same low risk.

5. Conclusion and Recommendations

From the results of various algorithms, we can highlight the following points:

Due to varying degrees of importance associated with the outage of various loads, it seems reasonable to have the units with low probability of outage serve the more important loads.

To avoid the high traffic in transmission lines, the risk of the transmission line outage may be considered, too.

Regarding network incidents, any sudden load surge on each bus may be considered as a network fault. This allows us to calculate the network risk of such situations.

For provision of the total network reserve, the consumer side reserve of some buses can be utilized. In other words, when a network fault occurs, the consumers of power may reduce their loads thereby help the system in restoring its balance.

Since the costs of risk for industrial and residential loads are different, the spinning reserve of a power network should be allocated in such a way that loads with high-cost- risk are protected against outage threats. We propose that in addition to probability of unit failure, future studies should take into consideration the probability of transmission line failures and the failures of high-voltage substations.

REFERENCE

[1] Z. Song, L. Goel and P. Wang, Optimal Spinning Reserve Allocation in Restructured Power Systems, IEEE Proceedings –Generation, Transmission, and Distribution, Vol. 152, No. 4, July 2005.
[2] Roy. Billinton, R N Allan, Reliability Evaluation of Power System, ISBN 0306453596, Springer, Jan 1996.
[3] Y. Hsu, S. Lee, et al, Operating Reserve and Reliability Analysis of the Taiwan Power System, IEEE proceedings, Vol. 137, pt. C, No. 5, Sep. 1990.
[4] IEEE Committee Report: "The IEEE Reliability Test System", Report 96 WM 326-9 PWRS, Presented at the IEEE Power Engineering Society Winter Meeting, Jan 1996, Baltimore, MD, USA.

Notes on Implementation of Novel Approach for Microscopic Measurement with Stereo Correlation

Avramov, M. Z., Soklev, B
University of Sofia,
Department of Physical Chemistry
mavramov@chem.uni-sofia.bg

Abstract-In the current presentation we would like to propose a method for collecting of stereovision data which can be extended to robust regression technique for multi-image stereo correlation of microscopic data.

3D Stereo vision has evolved in the last 30 years from a theory to a real application. Getting its motivation in one side from topography and space imagining and on the other from robotics the stereovision approach for recognition and measurements has been extensively studied and developed [1-3]. The purpose of this paper is to propose a stereo vision approach for micro and nano-measurements. Recently the interest on that area has been raised significantly and several publications appeared [4-14]. In this paper the authors would like to outline a method of collecting stereovision data and to point out on a clear opportunity for improvement of the collection of data which would eventually resolve multiple problems with ambiguities and occlusions in the stereo correlation algorithm.

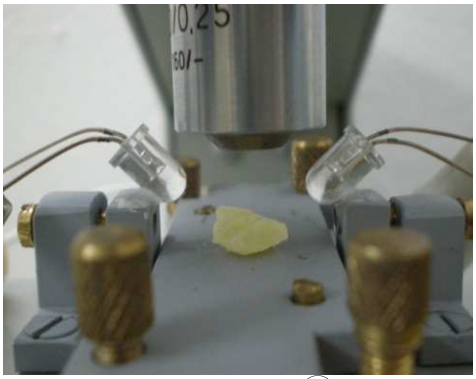
APPROACHES

Stereo vision is based on two images of an object taken from two slightly different view point. To achieve that it is possible to follow two approaches – watching the object from two different optical systems located near to each other or moving/rotating the object and taking both images from two (or more) positions of the object.

For the purpose of optical microscopy the first approach can be achieved using classic stereomicroscopy technique [4]. The second approach needs some special additions to be implemented. As the depth of view of classic microscopy is relatively small two options are needed – first a pinhole to increase the depth of view of the microscope and a special microscope stage which provide adjustable rotation to allow rotation against a defined axe in the specimen which lay in the focus plane of the microscope.

A stage which can be inclined in different angles is designed. The axe of rotation of the stage is placed higher then the stage plate so that the angle change does not defocus the specimen.

The actual stage allows adjustment of the bottom plate of the stage in a way that the specimen is positioned properly in the focus plane of the microscope in all angles.



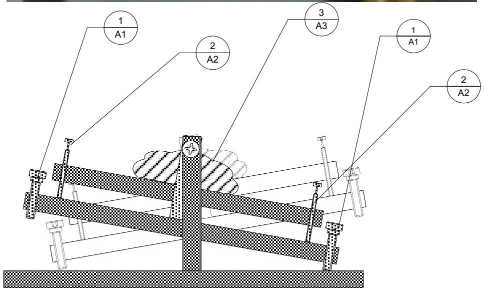


Fig 2. Micro-stage mechanical implementation and image. The stage allow regulation of the vertical position of the specimen (screws 2) and fixation of the angle on both directions (screws 1).

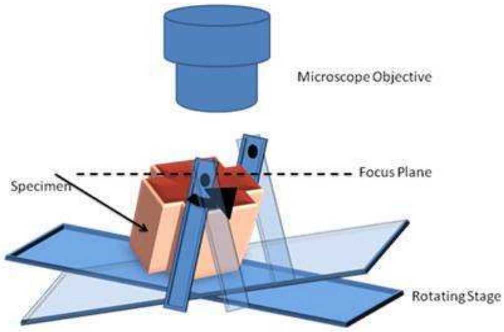


Fig 1. Scheme of single optics stereoscopic micro-measurement system.

The actual tolerance for the inclination angle is to be adjustable so different granularity of the surface can be observed without significant occlusions from stereo-correlation.

CALIBRATION OF A STEREO-MICRO-MEASUREMENT SYSTEM

Model for calibration should handle two important issues - Calibration of camera projection and distortion model and calibration towards epipolar constraints. Rectification process is important to reduce “false” matching points and potential occlusion or ambiguity issues. Nevertheless due to mechanical implementation of the angle changing stage it may lead to small tolerances in the calibration (Deviations of order of magnitude 1-2px over the range of 1280pxs may take place). The described mechanical mistake in angle definition is discussed in Resolution section. To avoid the problems involved by the mechanical imperfection a relaxed rectification approach is used. The matching is done in a rectangular area instead of strictly following epipolar lines (see Fig 3.) [17].

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_91, © Springer Science+Business Media B.V. 2010

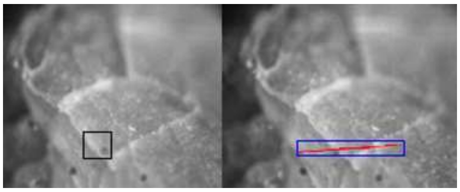


Fig 3. Correlation region (Black rectangle - left is the correlation BOX, Blue rectangle – right is the correlation area, red line – right is an example epipolar line).

RESOLUTION AND PRECISION

Horizontal resolution X, Y direction of the measurement is defined (practically) by the resolution of the used sensor. The Z direction (depth) resolution depends on sensor resolution and observation angle difference. Usually a sub-pixel estimation of the correlation is used in most implementations which may deliver 0.1 even 0.01 pixels according to some authors [9]. Significant impact on the precision of the measurement can be attached to the noise of the sensor. The disparity range obtained in angles of 10^{O} is of order of magnitude of up to 10 pixels. With the discussed above sub-pixel precision one can expect resolutions in the Z direction of 100 to 1000.

PRELIMINARY RESULTS

To outline the possibilities of the proposed measurement technique few experiments with optical materials obtained in the department of Physical chemistry [15] have been done.

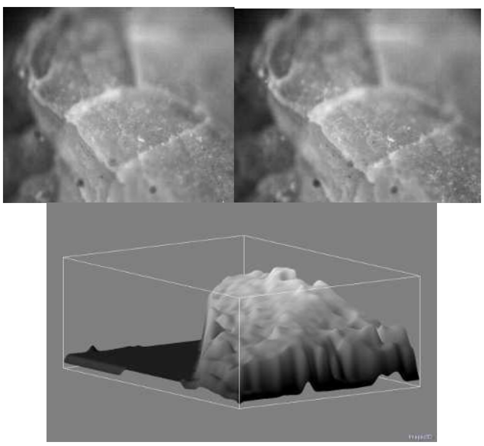


Fig4. Surface of SiO_2 dotted with Sm. Scanned and stereo-reconstructed with the stereo micro vision system.

Correlation of the images has been done with a custom assembled procedure based on comparison of gradients and phase of the images using Minimum of Squared Differences.

$$d = \sum_{square}\left[(1-\alpha)\left(\Delta I_x^{\;2} + \Delta I_y^{\;2}\right) + \alpha\left(\Delta I_x / \Delta I_y\right)^2\right]$$

Where $\Delta I_x = I_{i+1,j} - I_{i-1,j}$, $\Delta I_y = I_{i,j+1} - I_{i,j-1}$ and α is chosen to be 0.5.

Bidirectional matching for elimination of ambiguities has been applied. To avoid need of rectification, searching has been done in a rectangular area. Obtained results have been checked for epipolar constraints [17] (see Fig. 3). This approach practically combines stereo-correlation with epipolar calibration.

Another interesting option may be provided using SEM images of one specimen obtained from different angles of observation.

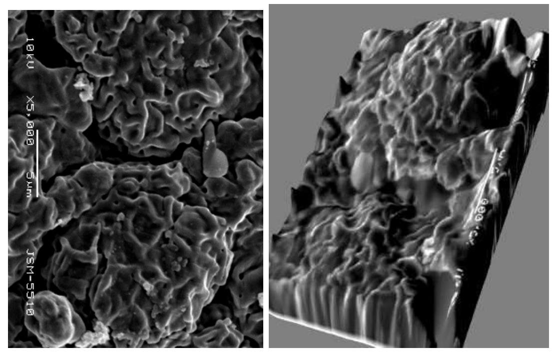


Fig 5. Left image and reconstruction of SEM of a Cu coated surface Stereo correlation from images with 15^{o} rotation of the stage magnification used is x5000.

FUTURE DEVELOPMENT

Once having observed a defined object from different points it sounds reasonable to extend the proposed approach to collect multiple images from many different angles. Following Radon transformation[18] approach used in Computed tomography one can imagine similar approach in light microscopy with tilting stage.

A simple device extension of the current stage is in development at the moment which will give the possibility to collect multiple images from different observation angles. For that purpose a precise synchronization of a mechanical actuator moving the stage is the main focus of current efforts.

ACKNOWLEDGMENT

This work is funded by National Science Foundation – Project NT1 / 01.

References

[1] Daniel Scharstein1 and Richard Szeliski, A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms November 2001,Technical Report MSR-TR-2001-81.

[2] Asim Bhatti, Saeid Nahavandi, Stereo Image Matching Using Wavelet Scale-Space Representation International Conference on Computer Graphics, Imaging and Visualisation (CGIV'06) pp. 267-272

[3] Frank Candocia and Malek Adjouadi A Similarity Measure for Stereo Feature Matching IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 6, NO. 10, OCTOBER 1997

[4] SCHREIER H. W. ; GARCIA D. ; SUTTON M. A. ; Experimental mechanics, 2004, vol. 44, n°3, pp. 278-288 Advances in light microscope stereo vision

[5] Kayaalp, A.E. Rao, A.R. Jain, R.; Scanning electron microscope based stereo analysis [for semiconductor IC inspection]; *Computer Vision and Pattern Recognition,* 1989. Proceedings CVPR '89., IEEE Computer Society Conference on;

[6] Li, N. (2007) Full-field Thermal Deformation Measurements in a Scanning Electron Microscope by 2D Digital Image Correlation. *Experimental Mechanics*

[7] Zhang, Dongsheng (2006) Displacement/strain measurements using an optical microscope and digital image correlation. *Optical Engineering* 45(3)

[8] Germaneau, A. (2007) 3D Strain Field Measurement by Correlation of Volume Images Using Scattered Light: Recording of Images and Choice of Marks. *Strain* 0(0)

[9] Sutton, Michael A. (2007) Three-dimensional digital image correlation to quantify deformation and crack-opening displacement in ductile aluminum under mixed-mode I/III loading. *Optical Engineering* 46(5)

[10] Gray, Mike H. B. (2006) Methodology for determining the influence of soft surface layers in porous solids on compressive behavior: Application to NASA's lightweight ceramic ablator. *Review of Scientific Instruments* 77(2)

[11] Sutton, M. A. (2007) Scanning Electron Microscopy for Quantitative Small and Large Deformation Measurements Part I: SEM Imaging at Magnifications from 200 to 10,000. *Experimental Mechanics*

[12] Scrivens, W. A. (2007) Development of Patterns for Digital Image Correlation Measurements at Reduced Length Scales. *Experimental Mechanics* 47(1)

[13] Germaneau, A. (2007) Full 3D Measurement of Strain Field by Scattered Light for Analysis of Structures. *Experimental Mechanics*

[14] Tiwari, V. (2007) Assessment of High Speed Imaging Systems for 2D and 3D Deformation Measurements: Methodology Development and Validation. *Experimental Mechanics*

[15] S.Gutzov, G.Ahmed, N. Petkova, E. Füglein, I. Petkov, Preparation and optical properties of samarium doped sol-gel materials, J Non-Crys Solids 354, 29, (2008) 3438-3442.

[16] T.Spassov, SEM microscopy of Cu Coated surface – Unpublished results.

[17] M. Avramov, Non-calibrated Stereo correlation algorithm for microscopy applications – Unpublished.

[18] Deans, Stanley R. (1983). *The Radon Transform and Some of Its Applications*. New York: John Wiley & Sons.

Distance Interaction in Education Processes using a Telepresence Tool

Fáber Giraldo, Ángela Jiménez[1], Helmuth Trefftz, Juliana Restrepo, Pedro Esteban[2]
Universidad del Quindío, Colombia
[1]{fdgiraldo, amjimenez}@uniquindio.edu.co,
EAFIT University, Colombia
[2]{htrefftz, jrestre, pesteban}@eafit.edu.co

***Abstract*- Distance interaction among educational communities is becoming increasingly important. Courses and talks are shared among institutions and individuals who are located in different places. The Telepresence tool, developed by Eafit and Universidad del Quindío, permits course sharing among the institutions. Instructors from one institution can teach students of the other one, thereby sharing the positive features of each course. In this paper, we present the recent improvements of the Telepresence tool. We also report on the use of the tool in a Multi-Variable Calculus course, designed in accordance to the Teaching for Understanding (TFU) pedagogical framework..**

Keywords: Telepresence, distance education, Multi-Variable calculus, Teaching for Understanding

I. Introduction

The Telepresence tool aims at supporting distance-education processes through the application of pedagogical concepts, such as TFU and the use of novel technologies (high-speed networks and shared virtual environments), in order to create teaching environments that allow the instructor and students to interact remotely.

Currently, most distance-education processes are mostly based on web-published contents. Interaction among the instructor and the students happens in an asynchronous manner, using forums or e-mail. We claim that this interaction is not appropriate to explain complex concepts. Experiences reported in [1], [2], [3], [4] and [5], show that the use of shared virtual environments with a Telepresence tool has fostered interesting interactions among the instructor and the students, that would be difficult to attain even in face-to-face education. Additionally, when students interact with virtual objects, the abstraction effort is reduced, thereby fostering the acquisition of knowledge and skills.

In a similar manner, designing the courses following TFU, allows the instructor to focus on the development of competences and skills in students, and this is a fundamental objective of current education systems. The pedagogical models used in distance-education are mostly based on traditional approaches. In other cases, the pedagogical approach is reduced to the interaction with a technological tool. The use of TFU allows the instructor to keep a log of each and every student's progress, and to focus on the development of competences and skills of the course at hand. These type of experiences that involve pedagogy and technology allow for a rapid incorporation of novel technologies, such as Virtual Reality and high-speed networks. Our experience with this project has been that successful practices can be shared between institutions when using the Telepresence tool. The ultimate goal of the project is to create and maintain a repository for creating and sharing learning objects among institutions.

II. The Telepresence tool

When using our Telepresence tool, instructors create a shared virtual environment that allows the instructor and the students to interact from remote places. The tool allows users to load graphic files to support teaching, and, more interestingly, to load a 3D interactive simulation of the concept at hand. The 3Dcontents are currently created using Java3D.

The most important enhancements of the Telepresence tool, implemented both by Eafit University and Universidad del Quindío, are presented in the next subsection.

A. Networks

Developers from Universidad del Quindío have implemented a multi-point communications protocol that allows one instructor to communicate with several groups located in different places. These changes were done by extending the previous network protocol. A new message server was added creating a local cache of the remote users.

B. Learning Objects

We have also implemented a program that packs each content as a SCORM learning object, based on the RELOAD Editor (http://www.reload.ac.uk/new/editor.html). The example of that object is shown on Figure 1. These contents can be sent to remote users, when the content is not stored locally.

After the content is sent, the tool will allow users to interact over the 3D scene.

C. Mathematical contents

Members of another Project at Eafit University have developed a 3D calculator. The calculator allows users to

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_92, © Springer Science+Business Media B.V. 2010

Fig. 1. Example of a SCORM-packaged content loaded in Telepresencia tool.

write the equation of a function involving X, Y and Z variables, and plots the corresponding surface. This software is written in C++. In order to add the functionality of the 3D calculator to our Project (written in Java), we are making use of JNI, a software tool that allows a Java program to call functions written in C. This allows the Java program to retrieve a mesh of points in space that will be later plotted in the Telepresence tool using Java and Java3D. Cartesian, polar, spherical and cylindrical coordinate systems can be used. This functionality was requested back in 2005, when the Telepresence tool was used in a course on Physics for Civil Construction professionals.

D. *Handling voice and video*

The current teleconferencing system is based on the "Java Media Framework" (JMF), which has not been updated since 2003. Since we are not aware of other java based applications for handling voice and video, we are considering the use of external tools. We are currently evaluating several platforms, including: jffmpeg, jvlc and FMJ. The first tool to evaluate will be JVLC, based on VLC. JVLC provides an API that can be used directly from the Telepresence application. VLC is, by itself, a video-stream server.

A new Graphical User Interface has been proposed (see Figure 2), it allows the instructor to switch easily among contents, without the need to close and re-open the application.

Lastly, since the development group is distributed geographically, we are using a code repository in GoogleCode in order to keep the current version.

III. INSTRUCTIONAL DESIGN

The quality of the contents that are included in the tool have the utmost importance for the group. Since TFU is used as

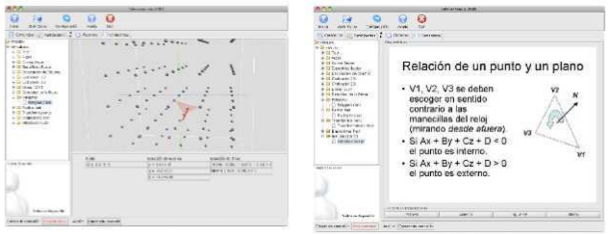


Fig. 2. GUI of the Telepresence tool.

pedagogical framework, all members of the research group received instruction in the basis of the framework.

A. *Teaching for Understanding - TFU*

Several pedagogical studies have explored methodologies to teach instructors adopt strategies that allow students to attain a better and more profound learning. One of these studies is the one led by researchers of Project Zero, at the Graduate School of Education at Harvard University, who have formulated the Teaching for Understanding (TFU) pedagogical framework. TFU proposes both instructors and students to participate actively in building and acquiring knowledge [9]. Ausubel, when creating the meaningful learning approach, suggests that learning implies a restructuring of ideas and concepts within the student's cognitive structure [7], [8], [9]. Ausubel also proposes the use of Conceptual Maps as a strategy to acquire new concepts. Concept Maps (CM) serve as organizers of previous knowledge and as a strategy to approach new concepts. Concept Maps can increase meaningful learning, helping the student to discover new relationships between the new concepts and those existing in his/her cognitive structure.

When a student is exposed to learning new concepts, knowledge has to be presented in a way that captures his/her attention. In this sense, Ausubel [7] provides the following criteria to structure teaching: (i) Knowledge to be learned has to be clear and explicit, presented in a language according to the student's previous level, (ii) The student has to have the previous relevant knowledge, (iii) the student has to choose to learn meaningfully, in other words, he/she has to be motivated towards learning. When a student learns, he/she builds meanings, i.e., organized cognitive structures that are related to current knowledge, thereby producing meaningful learning. When relationships are established in an arbitrary form, or are not generated at all, memorization learning takes place. Memorization learning is easily forgotten.

One of the strongest motivations that human beings have in order to attain a concept, is the possibility of applying it in the solution of existing problems. In order to attain this goal, the student must have an appropriate set of pedagogical tools. In TFU, the "portfolio" is used. The portfolio is a collection of activities and tangible resources that the student uses and builds during his/her learning process; it is the student's learning log.

The construction of a portfolio can be divided in three phases: (i) Concept Exploration (ii) Guided Research and (iii) Final Performance. In each of them, students show their work to their peers, the instructor and experts. The different phases are described next:

1) Concept Exploration. The student's environment is explored in relation with the topic at hand. In this phases the individual or group's interests become evident.

2) Guided Research: Each student or group, with the help of the instructor, selects a Project in which to apply the concepts of the concept at hand.

3) Final performance: Students present their findings and the solution they have implemented. They present the performances they have done during the project.

The portfolio is built by students and they use it either to revise concepts or cognitive relationships, or to form new ones.

The instructor's relationship with the student plays a very important role in the learning process. He/she has to favor the presentation of new knowledge in an orderly and appealing fashion, and should try to foster the student's interest on learning by applying teaching strategies that encourage learning.

TFU integrates various pedagogical theories [6] in order to Foster the students deep understanding of the proposed contents, particularly those that are fundamental when building the knowledge of the subject. For this purpose, it is important to link the students' interests with meaningful real-life situations. TFU does not propose a unique way to be applied, and for this reason the way how knowledge is presented to the student and the instructor's own understanding play a key role in the process.

B. The course

The way how we relate to our environment and the capacity to extract various properties from the objects in it, are key issues in the teaching-learning process. Based on these ideas, the course on Multi-Variable calculus was chosen, since it allows students create relationships between the environment and the mathematical concepts, as well as the concepts covered in other courses in the same semester. For this reason, the instruction modules to be used with the Telepresence tool were designed based on the syllabus of the Multi-Variable Calculus subject. Figure 3 presents the training process of TFU for teachers of Calculus in Universidad del Quindío

C. Course design under the TFU

Fig. 3. Training on TFU for the whole research team – June 2008.

The Multi-Variable calculus course is part of the Engineering Curriculum at the universities that take part of the Project. The design of the course was built around the question "How can I apply Calculus in my field of study?". With this question we try to encourage students to relate the course with other subjects in their curricula, and also with concrete objects in their environments. As part of the course, the students look for real-life objects that can be modeled with the equations covered in the course.

Calculus-related concepts that are particularly important are: vectors on the plane and vectors in space, partial and directional derivatives, multiple- and line-integrals. These concepts lead to the concept of "surface", therefore, students appreciate how the surface's concept is key for understand the Multi-Variable calculus course. In this course, students are asked to establish relationships between real-life objects and the concepts of the course. For this reason the concept of surface was defined as one of the key concepts.

For each concept in the course, several ways were proposed for students to demonstrate their deep understanding:

1) Concept Exploration. Students created an idea network, in which students put together the ideas they have about calculus and how it's related to their life. Concrete applications of calculus in various engineering fields were presented by the instructor. During this exploratory phase, students chose a problem to be solved by them. The problem was supposed to be related to their field of study.

2) Guided Research. Students starting solving the chosen problem, using the course contents and the technological tools to support the process. During this phase, students created a report with the problems they faced and the solutions they found

3) Final Performance. Each student demonstrated his/her understanding in a final Project that was presented to their peers.

IV. THE USE OF THE TELEPRESENCE TOOL IN THE MULTI-VARIABLE CALCULUS COURSE

One of the main objectives of this Project is to use the Colombian high-speed network (called RENATA). It links the research community in the country with the international research community, as a platform to run our Telepresence tool. The added value of the use of RENATA as communication platform is its power to Foster communication and collaboration among the various communities involved.

RENATA is the academic high-speed network in Colombia. It is linked to other academic high-speed networks in the world and allows the Colombian academic community to share data, information, lab equipment and transfer large amounts of data with other communities in our country or abroad. Also, RENATA enables the use distant of lab equipment and access to academic information between Universities in Colombia.

The Telepresence tool, running over RENATA, will allow an instructor from Eafit University to teach a Multi-Variable calculus to students in Universidad del Quindío. It will also allow an instructor from Universidad del Quindío to teach students at Eafit University. In both cases, the instructor and the students will interact in a shared virtual environment. The 3D content to be used are being built by the development groups.

CONCLUSIONS

The opportunity to integrate academic and technical groups from different universities, each with different perspective on the education process makes the collaborative work more enriching and interesting. Some of the highlights of the current collaborative work are:

1) Pedagogical training on TFU of the math instructors involved with the experience.
2) Training on the use of virtual resources by the instructors that take part of the design of the Multi-Variable Calculus course at both institutions.
3) Collaboration and integration of the two development groups in creating a stable single code base.

This integration will result on an improvement of the teaching process of the Multi-Variable Course at both universities, as well as a better utilization of the available technological resources by the students at both universities.

This project will foster the use of new technologies in the teaching-learning processes that are built on top of high-speed networks like RENATA. Our goal is to continue improving the Telepresence tool in order to make possible new and novel ways of interaction between the instructor and the students. High-speed networks should allow the academic communities to share learning objects and promote the integration and collaboration of academic communities in our country and abroad.

ACKNOWLEDGMENT

We express our gratitude to Eafit University, Universidad del Quindío and Colciencias (National Institute for the Advancement of Science in Colombia), for their support in the Project entitled “Evolution of a Telepresence tool and its application to distance learning in regions of the country using RENATA” (Code IF005-07).

The authors also express our gratitude to the students involved in the Project as research assistants:

EAFIT University:
- Alex Restrepo (alexrestrepo@mac.com)
- Camila Peñaloza (camilapb@gmail.com)

Universidad del Quindío:
- Julián Alejandro Lamprea (julianchoss@gmail.com)
- Jonny Alexander Vélez Calle (jonal_vc@yahoo.com)
- Daniel Andrés López (danilori24@hotmail.com)
- Daniel Andrés Peláez (daniel0788@hotmail.com)
- Julián Esteban Osorio (juesoslo@hotmail.com)
- Héctor Danilo Hurtado (hedahuol_98@hotmail.com)

REFERENCES

[1] A. Agudelo, L. Escobar, J. Restrepo, C. Zea, H. Trefftz. Telepresence for Distance Education: Lessons Learned. In proceedings of International Conference on Education (IADAT-e 2004) Bilbao, Spain, July 7 - 9, 2004.

[2] L. Escobar, E. Rendón, J. Restrepo, E. Montoya, C. Zea, H. Trefftz. Telepresence in Synchronous Distance Education Sessions. In proceedings of Second International Conference on Multimedia and ICTs in Education (M-ICTE 2003), pp 1070-1074, Badajoz, Spain, December 3-6, 2003.

[3] F. Giraldo, A. Jiménez P. Esteban, H. Trefftz. Collaborative Virtual Environments for Teaching Physics. In proceedings of International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering - CISSE 2006. - Institute of Electrical & Electronics Engineers (IEEE) University of Bridgeport December 4-14 2006.

[4] Orozco, Carlos Esteban, Esteban Duarte, Pedro Vicente, Trefftz Gómez, Helmuth. Collaborative and Distributed Augmented Reality in Teaching Multi-Variate Calculus. The Fifth IASTED International Conference on web-based education, Puerto Ballarta, México, 2006.

[5] H. Trefftz, P. Esteban, A. Quiroz, F. Giraldo, A. Jiménez, E. Villegas. Producto De Telepresencia para la Educación Superior en el Ámbito Nacional. Cuaderno 51 - Enero 2007. ISSN 1692 – 0694.

[6] Blythe, T., & et al (1999). La Enseñanza para la Comprensión. Guía para el docente. Paidós, Argentina.

[7] Novak, J. D., & Gowin, D. B. (1999). Aprendiendo a Aprender. Martínez Roca, Barcelona.

[8] Stone W, & Martha, C. (1999). La Enseñanza para la Comprensión. Vinculación entre la investigación y la práctica. Paidós, Argentina. Consultado en marzo de 2006.

[9] Díaz-Barriga F., & Hernández R.G. (2002). Estrategias Docentes para un Aprendizaje Significativo. Una Interpretación Constructivista. McGraw Hill, Segunda Edición, México.

A Study on the Lifetime Estimates of Aluminum Cable, Steel Reinforced (ACSR) Transmission Lines Under Flame Exposure

Young-Dal Kim[1], Jae-Myung Shim[1], Dae-Dong Lee[2], *Member IEEE*
[1]Dept. of Electrical Engineering, Hanbat Nat'l Univ., Dukmyung-Dong Yuseong-Gu, Daejeon, 305-719, KOREA
[2]Dept. of Electrical Engineering, Hanyang Univ., Haengdang-dong, Seongdong-gu, Seoul, 133-791, KOREA .

***Abstract*-To acquire basic data for lifetime estimation, tensile test, which is a type of mechanical test of ACSR, is required. By identifying the correlation of such tensile test results to aging and performing statistical processing, the lifetime can be estimated. Regression analysis is the most commonly used statistical processing method.**

In this study, we performed ACSR tensile test on sound wires, wires deteriorated by artificial flame, and deteriorated wires collected from forest fire sites. The limit of lifetime analysis was selected as the 20% loss point of rated tensile strength which is the standard set by Canada's Ontario Hydro Research

I. INTRODUCTION

In Korea, there are many cases in which ACSR have been installed as power transmission lines or overhead lines a long time ago; in this situation, interest in the remaining lifetime of cables is rising. [1, 2]

In addition, if power transmission lines are deteriorated by forest fire or resulting flames, the zinc plating on the zinc plated steel cables may be damaged or lost resulting in rapid corrosion which not only adversely affects the tensile strength of the steel cables but also it is likely that the property of the steel cable itself will be altered.

Since ACSR are installed in a variety of environments encompassing clean areas, contaminated areas, salt sea areas, salt contaminated areas etc. and vary in time after installation, there are tremendous difficulties in estimating the lifetime of cables deteriorated by flames while the extent of corrosion differs due to the influence of such environments. Thus. we assume that a wide variety of factors affect the lifetime of power transmission cables.

Since the ACSR cables are built from an outside layer of aluminum conductor and an inside core of zinc plated soft steel cable, the aluminum forms an oxidization layer due to atmospherically corrosion while being exposed to the atmosphere after installation, the mechanical tensile strength of aluminum was reported to decrease logarithmically while the decrement ratio of the zinc plated steel cable's tensile strength was reported to decrease exponentially since the corrosion proceeds while the zinc is lost. [3].

ACSR power transmission cables will be not only subject to loss of tensile strength due to corrosion and fatigue but also subject to rapid corrosion if exposed to flames which results in oxidation of the steel cable and therefore the tensile strength will be decreased. While the heat might increate the strength of the steel cable, the properties of the steel cable's material itself will be altered so that it loses the properties of a steel cable. In addition, its expansion rate, which is a measure of expansion upon being heated appropriately heated, increases so that the entire power transmission line expands;thus, wear or fatigue may accumulate due to frequent collision between conductors resulting in shorter lifetime of the power transmission line.

Therefore, in this paper, we attempt to estimate the lifetime of ACSR power transmission lines subject to flames based on the lifetime estimates from destructive testing performed together with nondestructive testing and by researching lifetime estimation using the tensile strength decrease rate which is a mechanical testing method forming the basis of lifetime analysis.

II. ACSR POWER TRANSMISSION LINES LIFETIME ANALYSIS USING REGRESSION

To estimate the lifetime of ACSR power transmission lines subject to flames as pursued by this research, the lifetime estimation of generic power transmission lines has to be reviewed first. It would be possible to attempt estimation of ACSR lifetime through regression analysis on data acquired through tension tests to analyze the relationship between aging and tension loss rate as a direct method. As an indirect method, eddy current testing method can be used. Such test results will be explored through the test results on Ontario Hydro.

Tensile strength testing was performed on each conductor sample collected from the power transmission line; the aging of sound cables was set to 0 and the aging to tensile strength loss ratio calculated relative to that of sound cables. Here, the lifetime estimation function according to the regression analysis results is as follows:

$$Y = e^{1.41+0.024x} \tag{1}$$

Lifetime analysis is also conducted using the torsion malleability test. This is done by doing the test on conductor samples taken from installed power transmission lines and

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_93, © Springer Science+Business Media B.V. 2010

sound cables. The lifetime is estimated from the resulting data of aging to number of torsions from multiple conductors based on the average value of malleability tests on the outer layer of steel cables. Here, the formula for lifetime estimates according to the torsion malleability test is as follows.

$$Y = 40 - e^{1.86+0.027x} \tag{2}$$

III. Lifetime Estimation of ACSR By Tensile Strength Analysis

We have studied the lifetime estimates of ACSR lines by applying regression analysis on aging, tensile strength loss rate and the number of torsions from Canada’s Ontario Hydro Research. Each result shows that there was a slight difference in the years of estimated lifetime. In addition, the results were not specific but presented in average lifetimes estimated by extracting samples from single layered structures.

Therefore, in this research, we intend to estimate the lifetime of ACSR lines exposed to flames by only using the tensile test results.

In this research, we will perform lifetime estimates by only using the tensile strength results. The tensile strength tests were performed on samples including sound cables, cables subject to artificial flames and cables recovered from forest fires.

A. Sampling and Experiment Method

To understand the description of samples, each is named as shown on Table 1.

Table 1. Remarks

NEW	Sound cable
PH	Cable deteriorated by forest fire flames
AF_00	Artificial fire contacted with sample
AF_10	Artificial flame of 10cm from sample
AF_20	Artificial flame of 20cm from sample

The sound cable called NEW is new cable purchased directly from the manufacturer’s factory; tensile strength test was performed on this cable without any treatment. The cable called PH is a collected sample which was deteriorated by forest fire; it refers to 10 specimens each collected at a distance of 10m from a removed cable of 100 m total length. The specimen called AF is a sound cable exposed to artificial flame; AF_00 refers to a condition in which the artificial flame and the sample were in direct contact ; AF_10 and AF_20 mean that the sample and the artificial flame were separated by 10[cm] and 20[cm],respectively.

The tensile strength of such samples was tested using a tensile strength tester. Fig.1 below shows the artificial flame test in progress; the flame’s size is ca. 20 [cm]. Fig.2 shows the tensile strength test; the test sample’s length is 50 [cm] the parts fixed on the jig are 5[cm] at both ends while the tension speed is 100[mm/min].

Fig. 1. Shape of Artificial fire equipment

Fig. 2. Equipment of mechanical test

B. Lifetime Estimate Using Aging to Tensile Load Loss Rate by Standard

A piece of cable recovered from a forest fire was installed 20 years ago. It was used as a sample for estimating the lifetime of ACSR power transmission lines exposed to flames. We analyzed the cable’s condition and conducted tensile strength and torsion tests to estimate the cable’s lifetime. Since these cables have no reference point, it is impossible to estimate their lifetime directly. Therefore, we compared and reviewed cables by applying identical mechanical experiments on sound cables and cables deteriorated by artificial flames.

In addition, the lifetime estimation function for cables was based on reference [4] which estimated the lifetime not according to flame but the years elapsed after installation and the environment of the installation area.

The regression analysis can be categorized into linear regression analysis and nonlinear regressing analysis, respectively. As in the case of Canada's Ontario Hydro Research, we assume that, in a diverse set of samples with aging reaching from 19 to 70 years, nonlinear analysis can derive an advantageous result and that there is no problem in analysis using Canada's nonlinear function's results under situations like those in Korea, we used nonlinear regression analysis in this research. Fig.3 shows the tensile strength loss ratio for each of the 5 types of samples. Here, NEW and AF cables each show the rate of tensile strength change under artificial flame exposure while for PH cable, the change of tensile strength according to forest fire flame is shown for 10 samples.

Fig.4 shows lifetime analysis using Formula (1) on tensile strength loss rate calculated from cable collected from forest fire exposed to flame. Here, the black part is the tensile strength change rate of 10 samples collected from a forest fire site; the tensile strength values in case of 20 years of use period are superimposed. The curves labeled from 10% to 90% show the results of Formula (1) in percentage. Here, to find the lifetime estimation point, the point where the average value of the test sample's tensile strength on the 30% curve meets the 80% RTS curve is the lifetime estimation point. The x coordinate of this point is the lifetime estimate in years.

In addition, (a)~(j) of Fig.5 are the lifetime estimation curves derived using the tensile strength change rates by distance from flame and flame exposure time when artificial flame is applied. The black rectangle symbol shows the tensile strength change rate for each sample; as shown roughly in Fig.3, even at the same flame exposure time, the tensile strength change rate is shown to differ according to distance from the flame. Here, from bottom to top, the curves are for sound cables, AF_20,AF-10 and AF_00. In addition, the tensile strength change rates are shown for 20 years for comparison with forest fire samples.

Lifetime from (h) of Fig.5 can be estimated as follows: 82 years for the bottom sample which is a piece of sound cable; 58 years for the sample exposed to flames 20[cm] away; and, while not accurate, 35 years for the sample exposed to flames 10[cm] away. For the sample subjected to artificial flames, making an estimate is not possible.

C. Analysis and Discussion of Test Results

The estimated lifetime of cables in years was analyzed by the nonlinear regression result of the rated tensile strength to each sample's tensile strength. Based on a lifetime of 82 years of sound cables, this lifetime analysis has been performed on collected cable samples replaced due to deterioration by forest fire and samples which were exposed to artificial flames according to time.

The cable used as samples was collected after subject to oxidation fire; as shown on Fig.4, differences were found depending on the place of collection. We assume that this is because the forest fire did not occur on one location but moved by wind or firewood thus affecting cables differently; as shown in this figure, the average lifetime can be estimated to 65 years.

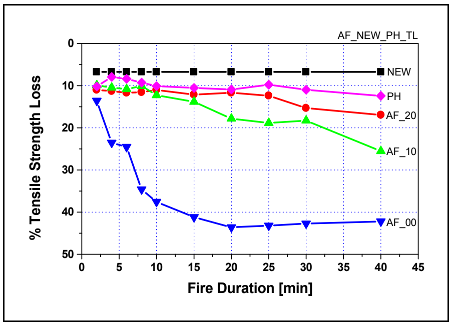


Fig. 3. Loss of tensile strength

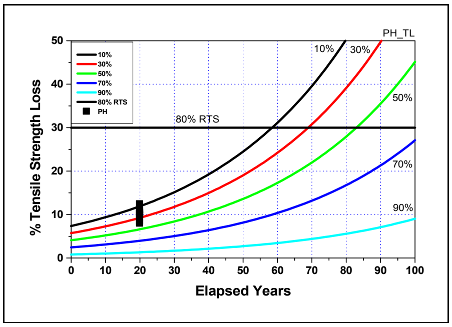
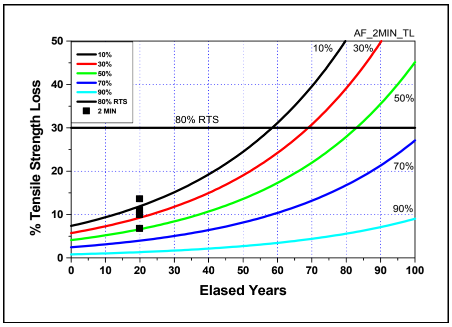
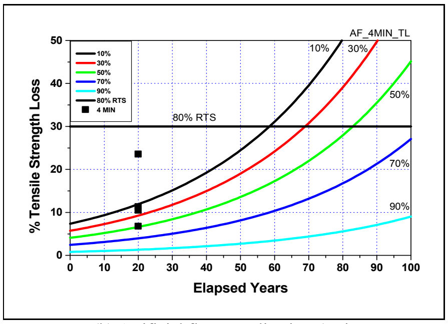


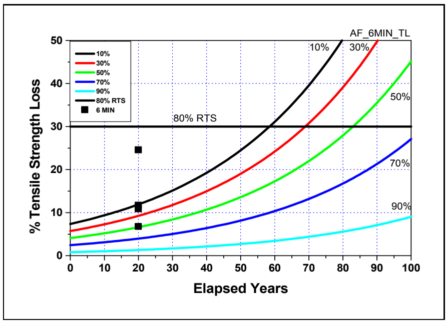
Fig. 4. Tensile strength loss for life-time



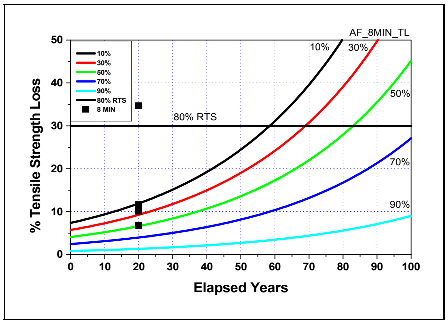
(a) Artificial flame application-2 min



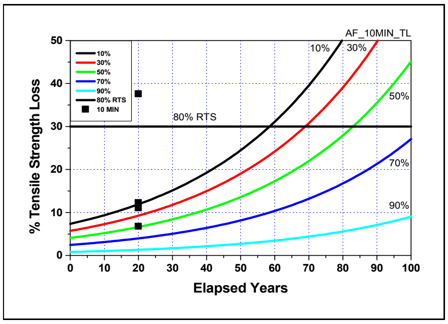
(b) Artificial flame application-4 min



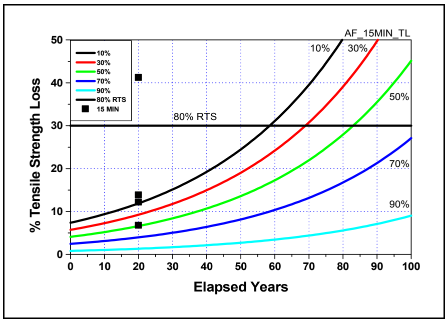
(c) Artificial flame application-6 min



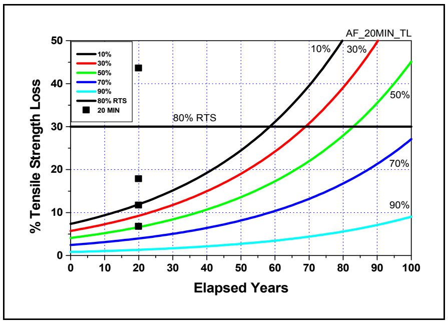
(d) Artificial flame application-8 min



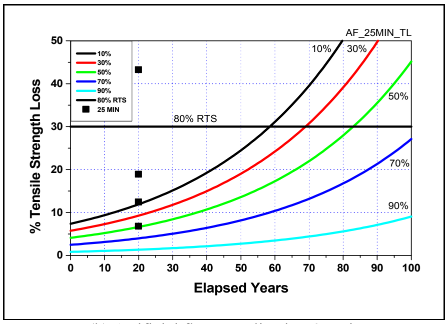
(e) Artificial flame application-10 min



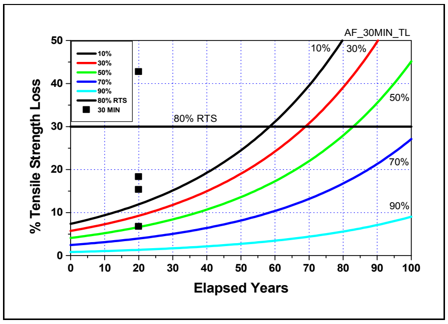
(f) Artificial flame application-15 min



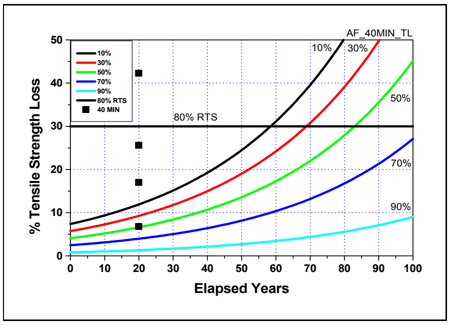
(g) Artificial flame application-20 min



(h) Artificial flame application-25 min



(i) Artificial flame application-30 min



(j) Artificial flame application-40 min

Fig. 5. Tensile strength loss for life-time Estimation of AF

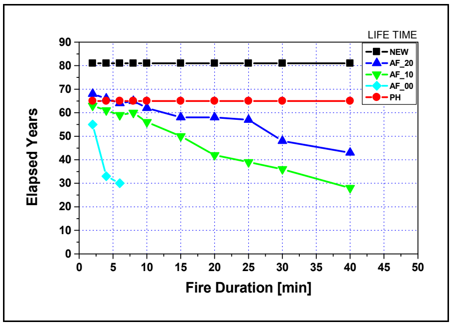


Fig. 6. Life-time Estimation

In addition, the lifetime estimation curve by distance and application time on cables with artificial flame applied is shown in detail on Fig.5; we see that, the cable lifetime decreases not only with longer exposure time to artificial flame but also with shorter distance to the flame. We assume that this phenomenon occurs since the zinc plating on the steel cable is lost due to flame; therefore, not only does corrosion occur which results in the steel cable's loss of tensile strength but also is the property of the steel cable material changed therefore affecting the loss of tensile strength. Therefore, if such cable sustaining damage from flame is not replaced but used for a long time remaining installed, we presume that rapid corrosion will occur on the steel cable thus resulting in severe loss of tensile strength. In addition, the loss will be larger on installation sites such of coastal or contaminated areas.

Therefore, in order to compare such conditions in a concrete manner, we summarized the lifetime estimates for sound cables, cables collected from forest fires, and cables with artificial flames applied, as shown on Fig.6.

If there are cables exposed to flames in the future, the replacement status of cables should be determined by estimating the changes in tensile strength loss rate as shown by the result of such experiments. This will enable a more stable and reliable power supply.

Finally, to increase the reliability of the nonlinear regression analysis used here, more samples and aging are required; therefore, we assume that, by continuous mechanical testing of cables in the future, we will be able to achieve more reliable results.

In collecting field samples, the selection of parts for collection is deemed important together with keeping storage of new cables at the time of installation and the recordkeeping of cable history since systematical sampling is required. In addition, the installation environments should be categorized into clean areas, salt water areas, salt contaminated areas, industrial areas etc. to facilitate more comprehensive studies.

IV. CONCLUSION

The purpose of this research was to estimate the lifetime of ACSR power transmission cables deteriorated by forest fire; for lifetime estimation on general aging , the formula derived from research of Canada's Ontario Hydro Research was used. Bu applying the separation distance between flame and cable and the tensile strength change rate varying according to flame exposure time, the lifetime varying by flame extent was estimated.

In particular, to estimate the lifetime of cables deteriorated by forest fire using the results of this research, the tensile strength's rate of change for each case was tested using an artificial flame generator; by comparing the two cases to each other, we acquired comparative data from which the extent of deterioration on cables deteriorated by forest fire can be estimated.

Therefore, the key of this research is the tensile strength change rate of each case; the comparative results for lifetime estimation are as on Section 3 above. By acquiring experimental data for a variety of conditions in the future and collecting such data on a database, we will be able to estimate lifetime by simply collecting cables deteriorated by forest fire and checking its tensile strength change rate.

REFERENCES

[1] Horrocks.,DJ., and Amm, D.E. "A Long Term Transmission Line Refurbishment Program", canadian Electrical Association, Spring 1987 Meeting, Vancouver.

[2] Harvard, D.G., et al. "Preliminary Studies for an Overhead Line Refurbishment Program at Ontario Hydro", Report to CIGRE WG22-211, Sarajevo, Jugoslavia, 1989.

[3] P.H. Schwabe & D. Pike, "The Measurement of Corrosion in Overhead Power Lines", ANTI-Corrosion, July 1988.

[4] The Math Works, Inc, Statistics Toolbox User's Guide, Jan. 1997.

A Study on the Vibration Phenomenon of a Power Transformer in Operation (154kV/60MVA/Single-phase)

Young-Dal Kim[1], Jae-Myung Shim[1], Woo-Yong Park[1], Sung-joong Kim[1], Dae-Dong Lee[2], *Member IEEE*
[1]Dept. of Electrical Engineering, Hanbat Nat'l Univ., Dukmyung-Dong Yuseong-Gu, Daejeon, 305-719, KOREA
[2]Dept. of Electrical Engineering, Hanyang Univ., Haengdang-dong, Seongdong-gu, Seoul, 133-791, KOREA .

Abstract - **Most of the causes of the failures of power transformers in operations are mechanical defects due to vibration. To identify these causes, the vibration phenomenon was measured from the power transformers in operation in the substations under the Daejeon Power Transmission District Office of KEPCO. The measurement was performed in a 6x4 structure on one side of the transformer. The vibration of the measured points was presented in 2D and 3D, and analyzed. The results according to the period in use of the transformer and the load applied to the transformer were analyzed. These results will be used as basic data to establish the vibration standard for the power transformer.**

I. INTRODUCTION

In 1842, James Joule explained the magnetostrictive phenomenon, which is a property of a material that causes it to change its shape when it is subjected to a magnetic field [1]. A device with a mechanical exciting force is accompanied by resonance when the external vibration frequency is equal to the natural frequency. Changing the natural frequency of the device by attaching a heavy object to it to change its mass or by connecting a differently shaped object such as a T-structure to it can be a measure to endure the specific vibration frequency [2]. An electric frequency of multiples of 120 Hz generated by the magnetostrictive phenomenon of the inner iron core is emitted outside through the transformer enclosure. At this time, the resonance is increased by the frequency equal to the magnetostrictive phenomenon frequency of the transformer enclosure [3-4].

This resonance can cause mechanical defects, but there are few data on the vibration of the power transformer in operation in Korea. To solve this problem, the vibration of the single-phase 154kV transformers in operation was measured so that the analysis results can be used as basic data for the vibration standard for the power transformer.

II. MEASUREMENT AND ANALYSIS OF THE VIBRATION OF THE TRANSFORMER IN OPERATION

In this study, the vibration level of a power transformer in operation was measured in compliance with the KS A 0603 standard (1998 Vibration Level Measurement) and KS B 07114 (2001 Mechanical Vibration and Impact: Mechanical Installation of the Accelerometer). A RION PV-95 accelerometer, a RION UV-06A charge amplifier, and a B&K pulse data acquisition/analysis device were used in the measurement.

The vibration measurement system was constructed as shown in Fig. 1 [5]. For the noise of the transformer, the audio frequency range for a human being is 20Hz – 15 kHz. In particular, the physical strength of the sound is determined by the sound pressure, but the strength felt by a human being varies also according to the sound frequency. The measured frequency range was set at 0Hz - 3.2 kHz (Step = 1 Hz).

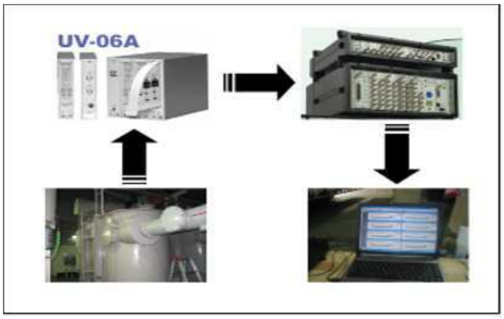


Fig. 1. Vibration measurement system

Thirty-two transformers were selected for measurement according to the period of their use: 15 transformers with a period in use of less than 10 years; 12 transformers with a period in use of 10 years or more but less than 20 years; and five transformers with a period in use of 20 years or more but less than 30 years. The 32 transformers in use were installed in nine substations. The operation conditions, installation conditions, and load capacity at the time of the measurement were also used in the data analysis. One hundred twenty vibration data were acquired at intervals of 0.5 s for 1 minute, and their trend was observed. Fig. 2 shows the instantaneous vibration data for Transformer No. 1 in Substation E, and Fig. 3 shows the trend of the vibration data for Transformer No. 1

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_94, © Springer Science+Business Media B.V. 2010

in Substation I for 1 minute, And Fig. 4 shows the instantaneous vibration data for Transformer No. 1 in Substation I, and Fig. 5 shows the trend of the vibration data for Transformer No. 1 in Substation I for 1 minute.

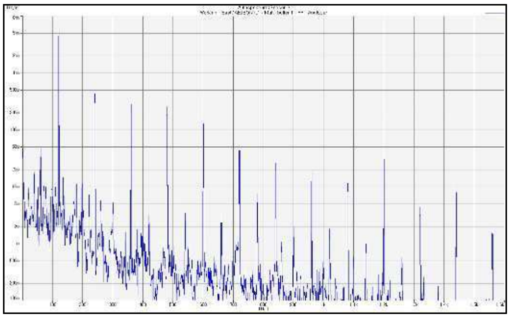


Fig. 2. Vibration data for Position No. 1 of Transformer No. 1 in Substation E

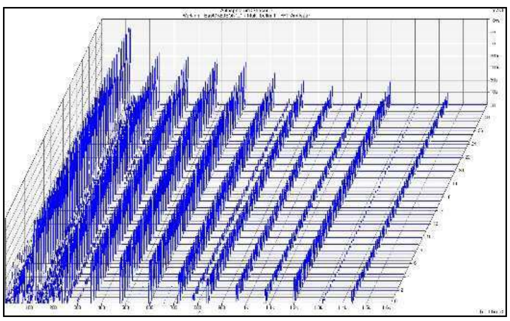


Fig. 3. Vibration trend at Position No. 1 of Transformer No. 1 in Substation E for 1 minute

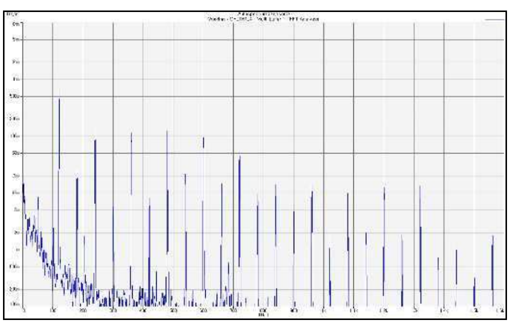


Fig. 4. Vibration data for Position No. 6 of Transformer No. 1 in Substation I

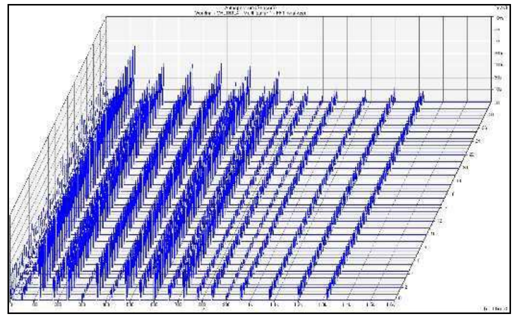


Fig. 5. Vibration trend at Position No. 6 of Transformer No. 1 in Substation I for 1 minute

The results of the vibration level measurement for the first transformer in each substation are shown in Fig. 6. Table 1 shows the peak value of the vibration level according to the frequency of the first transformer in each substation. As shown in Fig. 6, the transformer in Substation E generated the largest vibration, whereas the transformer in Substation I generated the smallest vibration. This result showed that a longer period in use generates a larger vibration. This is deemed to have been not only because the transformers had defects, but also because the transformer was not sufficiently supported by the old auxiliary equipment.

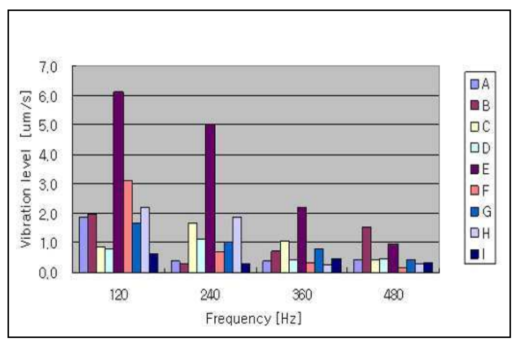


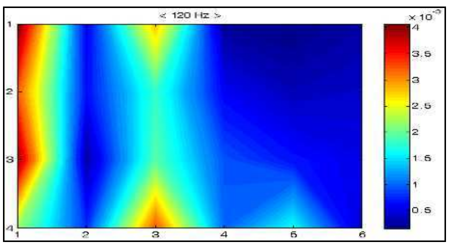
Fig. 6. Vibration levels of the No. 1 transformers

Fig. 7 shows the vibration map that exhibits the vibration on the side of Transformer No. 1 in Substation E, wherein the highest vibration level was measured at the rear left side for a frequency of 120 Hz. Fig. 8 shows the vibration map for Transformer No. 1 in Substation I. Fig. 7 shows the overall vibration of a transformer with a long period in use, and Fig. 8 shows the partial small vibration of a transformer with a short period in use.

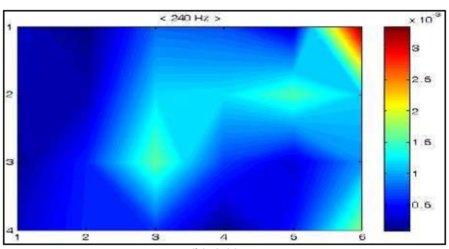
Table 1. Vibration Levels of the No. 1 Transformers by Frequency

(Unit: mm/s)

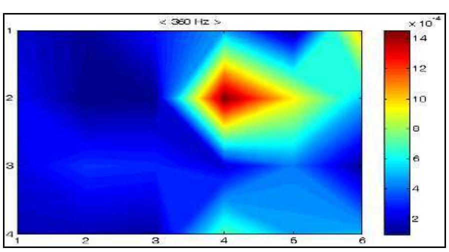
	120 Hz	240 Hz	360 Hz	480 Hz
A	1.87	0.38	0.39	0.44
B	1.95	0.29	0.70	1.53
C	0.86	1.65	1.07	0.44
D	0.81	1.15	0.43	0.45
E	6.12	5.00	2.19	0.97
F	3.08	0.67	0.33	0.16
G	1.64	1.04	0.79	0.44
H	2.23	1.86	0.25	0.29
I	0.62	0.28	0.46	0.32



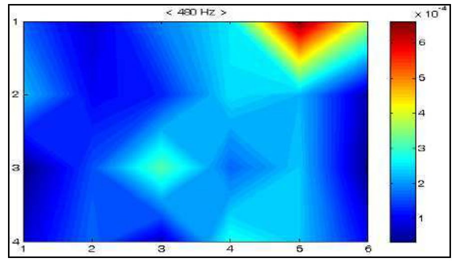
(a) 120Hz



(b) 240Hz

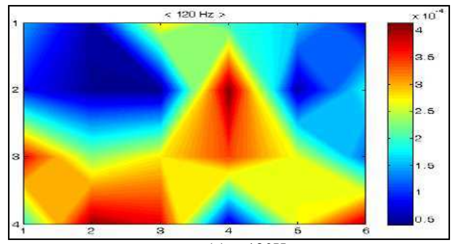


(c) 360Hz

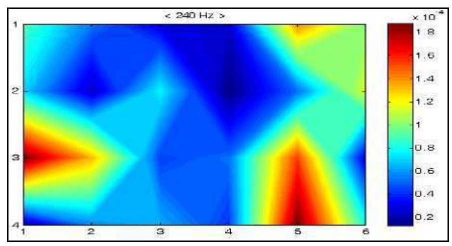


(d) 480Hz

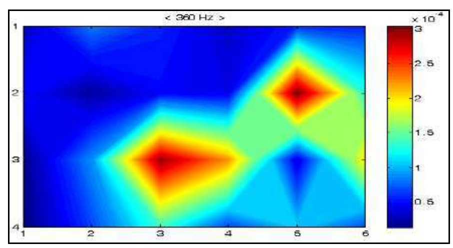
Fig. 7. Vibration of Transformer No. 1 in Substation E



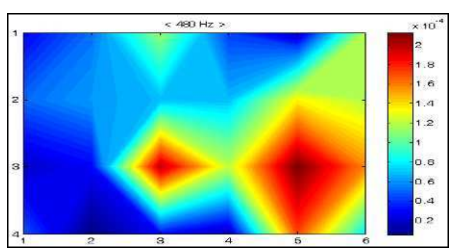
(a) 120Hz



(b) 240Hz



(c) 360Hz



(d) 480Hz

Fig. 8. Vibration of Transformer No. 1 in Substation I

The vibration level according to the manufactured year and by manufacturer was also analyzed. Fig. 9 shows an example for Manufacturer A, wherein the vibration level clearly increases according to the increasing period in use for 120 Hz but not for 240 Hz, 360 Hz, and 480 Hz. The example for Manufacturer B in Fig. 10 shows almost no change in the vibration level according to the period in use, which is deemed to have been because the transformer was designed in sufficient consideration of the effect of vibration. The example for Manufacturer C in Fig. 11 shows a high vibration level for the transformer with a long period in use, and on the contrary, recently installed transformers that show a high vibration level.

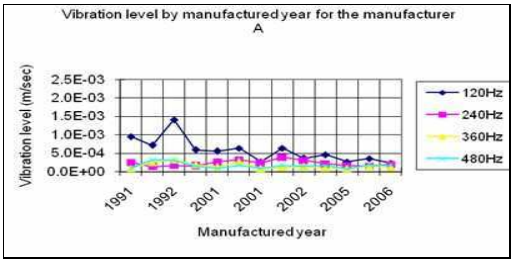


Fig. 9. Vibration level by manufactured year for Manufacturer A

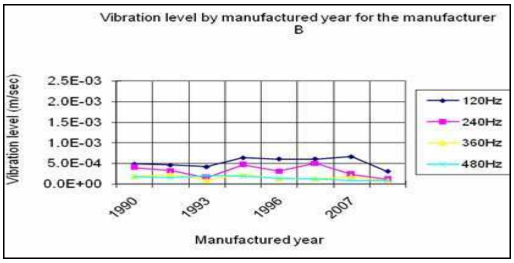


Fig. 10. Vibration level by manufactured year for Manufacturer B

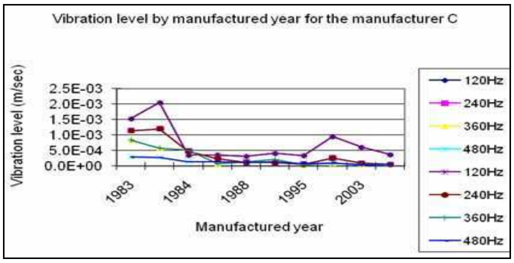


Fig. 11. Vibration level by manufactured year for Manufacturer C

The observed fluctuation of the vibration level under different loads is shown in Fig. 12. The vibration level did not change much vis-a-vis the load fluctuation, except for some transformers wherein the change was large. Therefore, it is deemed to have been due to the installation condition or other factors.

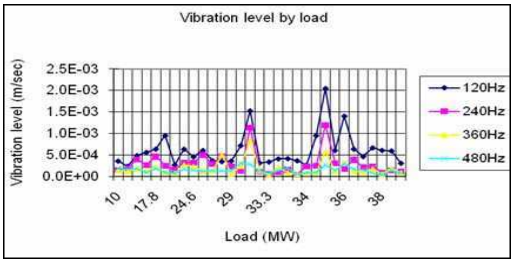


Fig. 12. Vibration level by load

III. CONCLUSION

Changes in the vibration level according to the period in use and the load were observed in single-phase power transformers in operation. All the transformers showed high vibration levels at a frequency of 120 Hz. There was a high risk of resonance at 120 Hz for the transformer enclosure, but it did not appear to have been significant. It is deemed to have been largely due to the damping of the structure. Some transformers in Substation E showed vibration levels higher than those of the other transformers, although there was no concern for the resonance phenomenon. It seems that one of the reasons why the resonance phenomenon did not increase was the resonance-resistant design of the transformer. Standards for transformer vibration measurement and evaluation can be established in the future based on the measurement and analysis of the vibration level of the transformers in this study.

REFERNCES

[1] Mou, Gang. Modeling and Control of a Magnetostrictive System for High-precision Actuation at a Particular Frequency. 2002. Pp. 4-5.
[2] Dukkipati, Rao V. Advanced Mechanical Vibration.
[3] David Jiles, "Introduction to magnetism and magnetic materials", Chapman and Hall, pp89-106, 1991
[4] M.J. Dapino, R.C. Smith, A.B. Flatau "An active and structural strain model for magnetostrictive transducers", SPIE symposium on Smart Structures and Materials, 1998, Paper#3329-24
[5] Ji Shengchang; Shan Ping; Li Yanming; Xu Dake; Cao Junling " The vibration measuring system for monitoring core and winding condition of power transformer" Electrical Insulating Materials, 2001. (ISEIM 2001). Proceedings of 2001 International Symposium on 19-22 Nov. 2001 Page(s):849 – 852

Distinguishing Fall Activities using Human Shape Characteristics

Homa Foroughi[1], Mohamad Alishahi[2], Hamidreza Pourreza[1], Maryam Shahinfar[3]
[1]*Department of Computer Engineering, Ferdowsi University of Mashhad, Iran*
[2]*Islamic Azad University, Mashhad Branch, Iran*
[3]*Islamic Azad University, Ramsar Branch, Iran*
ho_fo46@stu-mail.um.ac.ir, alishahi@gmail.com, hpourreza@um.ac.ir, m.shahinfar@yahoo.com

Abstract- **Video Surveillance is an omnipresent topic when it comes to enhancing security and safety in the intelligent home environments. In this paper we propose a novel method to detect various posture-based events in a typical elderly monitoring application in a home surveillance scenario. These events include normal daily life activities, abnormal behaviors and unusual events. Due to the fact that falling and its physical-psychological consequences in the elderly are a major health hazard, we monitor human activities with a particular interest to the problem of fall detection. Combination of best-fit approximated ellipse around the human body, horizontal and vertical velocities of movement and temporal changes of centroid point, would provide a useful cue for detection of different behaviors. Extracted feature vectors are finally fed to a fuzzy multiclass support vector machine for precise classification of motions and determination of a fall event. Reliable recognition rate of experimental results underlines satisfactory performance of our system.**

I. INTRODUCTION

Intelligent video surveillance systems are receiving a great deal of interest especially in the fields of personal security and assistance. The fundamental capabilities of a visual surveillance system are to provide recognition trajectories of the moving objects in the scene, and to use them for further processing. In the past few decades, vision-based surveillance has been extensively applied on industrial inspection, traffic control, security systems, and medical and scientific research. One of the application areas of video surveillance is monitoring the safety of elderly in home environments.

Human society is experiencing tremendous demographic changes in aging since the turn of the 20th century. And in most of our countries elderly people represent the fastest growing segment of the population and this trend will increase over the next years. According to a report of U.S. Census Bureau, there will be a 210% increase in the population with age of 65 and over within the next 50 years [1]. In the case of elderly people living on their own, there is a particular need for monitoring their behavior, such as a fall, or a long period of inactivity. A fall incident not also causes many disabling fractures but also has dramatic psychological consequences that reduce the independence of the person. It is shown in [11] that 28-34% elderly people in the community experience at least one fall every year, and 40-60% of the falls lead to injury. Thus with the population growing older and increasing number of people living alone, supportive home environments able to automatically monitor human activities are likely to widespread due to their promising ability of helping elderly people. Nowadays, the usual solution to detect falls is wearable sensors. These autonomous sensors are usually attached under the armpit, around the wrist, behind the ear's lobe or at the waist. These devices integrate accelerometer and/or inclinometer sensors and can monitor velocity and acceleration, vertical posture toward lying posture. However the problem of such detectors is that older people often forget to wear them, indeed their efficiency relies on the person's ability and willingness to wear them. Moreover in the case of noncontact sensors, they often provide fairly crude data that's difficult to interpret. Computer vision systems track the resident using cameras installed at vantage locations and attempt to detect a fall event based on image-processing algorithms that are designed to identify unusual inactivity [3].

In this paper we present a novel video analysis based approach for monitoring human activities with a particular interest to the problem of fall detection. The reminder of the paper is organized as follows: in section II we briefly review some existing vision-based fall detection systems, in section III our proposed system is introduced and more technical details are described. Experimental results are represented in section IV. Section V draws conclusions and outlines future work directions.

II. RELATED WORK

Recently some research has been done to detect falls using image processing techniques. A simple method was used in [5], [6] based on analyzing aspect ratio of the moving object's bounding box. This method could be inaccurate, depending on the relative position of the person, camera, and perhaps occluding objects. The works in [2], [7] used the normalized vertical and horizontal projection of segmented object as feature vectors. To overcome occluding objects problem, some

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_95,

researchers have mounted the camera on the ceiling: Lee [8] detected a fall by analyzing the shape and 2D velocity of the person. Nait-Charif [9] tracked the person using an ellipse and inferring falling incident when target person is detected as inactive outside normal zones of inactivity like chairs or sofas. Rougier [4] used wall-mounted cameras to cover large areas and falls were detected using motion history image and human shape variation. Other systems used the audio information or using 3D trajectory and speed of head to infer events [11]. These mechanisms tend to be more complex and need more additional cost. Despite the considerable achievements that has accomplished on this field in the recent years, there are still some clear challenges to overcome. To address limitations, this paper presents a novel approach which aims not only to detect and record fall events, but also other postures in a home environment with a considerable recognition rate. The main contributions of this paper are as follows:

- One of the major challenges of vision-based systems is the perceived intrusion of privacy because of the way that image data are transmitted and analyzed. To circumvent intrusion of privacy problem, the captured images in the proposed method are immediately filtered at the first level into motion features, which encapsulate only the motion vectors of the subject. Visual images are not stored or transmitted at any stage of the process so that it is impossible to reconstruct the abstract information back into the original image.
- Visual fall detection is inherently prone to high levels of false positive as what appears to be a fall might not be a fall, but a deliberate movement towards the ground. In other words, most of current systems [6], [7], [9], [10] are unable to discriminate between real fall incident and an event when the person is simply lying or sitting down abruptly. This paper presents an approach which is not inherently prone to false positives. This is because it incorporates an efficient classifier in assessing the presence of fall events.
- To simulate real life situations, we considered comprehensive and various movement scenarios consisting *normal* daily life activities such as walking, running, bending down, sitting down and lying down, some *abnormal* behaviors like limping or stumbling, and also *unusual* events like falling.

III. PROPOSED SYSTEM

This paper proposes a novel method for monitoring human posture-based events in a home environment with focus on detecting three types of fall. Our approach exploits computer vision techniques to detect and track people inside a single room with a single camera for classifying different postures and detecting falls. Since we are interested in analyzing the motion occurring in a given window of time, firstly we need to obtain the segmentation of moving objects. Therefore, a background estimation procedure is performed to separate motion from background. After the silhouettes are acquired, the next step involves extracting features and tracking their patterns over time. Because of the dependence of classification results on efficient representation of images by the feature vectors, this procedure is a determinant factor in the performance. Since changes in the human shape can discriminate if the detected motion is normal (e.g. the person walks or sits) or abnormal (e.g. the person falls), we analyze the shape changes of the extracted silhouettes in the video sequence. It seems that combination of approximated ellipse around the human body, horizontal and vertical velocities and temporal changes of centroid of human body; would provide a useful cue for detecting different behaviors. The extracted features will then be fed to the fuzzy multi-class support vector machine for precise classification of motions.

A. Foreground Segmentation

Background subtraction is a particularly popular method for motion segmentation and attempts to detect moving regions in an image by differencing between current image and a reference background image in a pixel-by-pixel fashion. In order to human segmentation and extracting the moving objects; here we use a background subtraction method described in [15], which is fairly robust and gives appropriate results on image sequences with shadows and highlights.

B. Feature Extraction

One of the issues of major importance in a recognition system is feature extraction, i.e., the transition from the initial data space to a feature space that will make the recognition problem more tractable. Intuitively, the human shape appears to be a good feature to look at as it captures the motion of most of the body parts. So we analyze the shape changes of the extracted silhouettes in the video sequence. To this aim, three main features that retain the motion information of the actions and can be easily obtained are selected to form the feature vector.

1) Best-fit Ellipse

When a motion is detected, an analysis on the moving object is performed to detect a change in the human shape, more precisely in orientation and proportion. The person is then approximated by an ellipse using moments [4]. An ellipse is defined by its center $(\bar{x}, \bar{y})$, its orientation θ and the length a and b of its major and minor semi-axes [13]. The authors believe that this representation is rich enough to support recognition of different events. For a continuous image $f(x, y)$, the moments are given by:

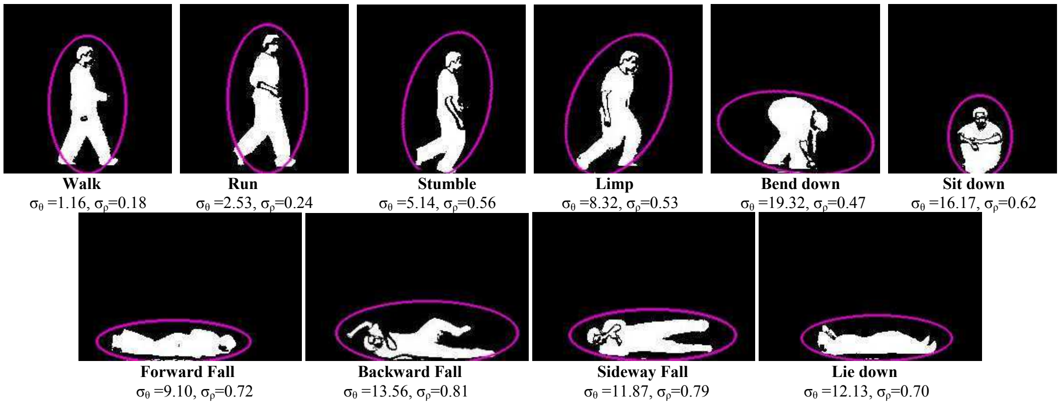


Fig. 1. Approximated Ellipsis of Different Human Postures

$$m_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^p y^q f(x,y)\,dxdy \quad \text{for } p,q = 0,1,2,... \tag{1}$$

The center of the ellipse is obtained by computing the coordinates of the center of mass with the first and zero order spatial moments. Then the centroid $(\bar{x},\bar{y})$ is used to compute the central moment as follows:

$$\mu_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\bar{x})^p (y-\bar{y})^q f(x,y)\,d(x-\bar{x})\,d(y-\bar{y}) \tag{2}$$

The angle between the major axis of the person and the horizontal axis x gives the orientation of the ellipse, and can be computed with the central moments of second order:

$$\theta = \frac{1}{2}\arctan\left(\frac{2\mu_{11}}{\mu_{20}-\mu_{02}}\right) \tag{3}$$

Values I_{min} and I_{max}, (the least and greatest moments of inertia) are computed by evaluating the eigenvalues of the covariance matrix [10]:

$$J = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix} \tag{4}$$

Then the major semi-axis a and the minor semi-axis b of the best fitting ellipse are given by [6]:

$$a = (4/\pi)^{1/4}\left[\frac{I_{\max}^3}{I_{\min}}\right]^{1/8} \quad b = (4/\pi)^{1/4}\left[\frac{I_{\min}^3}{I_{\max}}\right]^{1/8} \tag{5}$$

With a and b, we can also defined the ratio of the ellipse $\rho = a/b$ (related to its eccentricity). Fig. 1 shows examples of the approximated ellipse of the person in different situations. In order to detecting changes in the human shape discriminate an abnormal behavior from other normal activities, two values are computed:

- **The orientation standard deviation σ_θ of the ellipse**: If a person falls perpendicularly to the camera optical axis, then the orientation will change significantly and σ_θ will be high. If the person just walks, σ_θ will be low.
- **The ratio standard deviation σ_ρ of the ellipse**: If a person falls parallelly to the camera optical axis, then the ratio will change significantly and σ_ρ will be high. If the person just walks, σ_ρ will be low.

Note that σ_θ and σ_ρ are computed for a 2 second duration (approximately 60 frames). As Fig. 1 demonstrates, the orientation standard deviation of the ellipse and also the ratio standard deviation of the ellipse are completely distinctive for different postures.

2) Vertical/Horizontal Velocity

The velocity in the vertical direction (V_v) and in horizontal plane (V_h) are calculated and then two parameters were determined: peak velocities, and lead time of fall activities, defined as the time between both velocities exceeding the peak velocity of the normal activity (e.g., 1 m/s) and the time when the vertical position of the body reached either the height of the mattress or minimum.

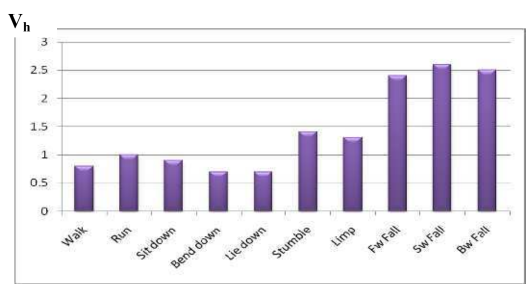


Fig. 2. Absolute Peak Velocity in Horizontal and Direction

There are two main observations for all the normal activities. Firstly, peak V_h and V_v are limited within 1m/s range, as shown in Fig. 1. Secondly, it was found that these peak velocities did not occur at the same time. However, for all fall activities, in contrast to the normal activities, the peak V_h and V_v tended to occur simultaneously. Besides, both V_h and V_v were found to exceed the 1 m/s range at about 300-400 ms before fall.

V_v

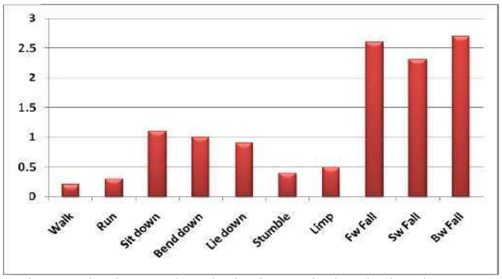
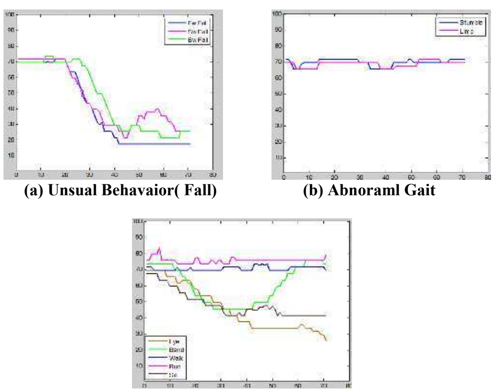


Fig. 3. Absolute Peak Velocity in Vertical and Direction

3) Temporal Changes of Centroid

Centroid is an effective way to estimate object motion in space. We choose to track the centroid because it is a potentially powerful and intuitive pointing cue if it can be obtained accurately, and has a large movement during a fall. Our proposed method tries to estimate a person's centroid in every frame of a video sequence and further infer his/her attentive behavior.



(a) Unusual Behavaior(Fall) **(b) Abnoraml Gait**

(c) Normal Daily Activity

Fig. 4. Centroid Displacement

Centroids can be computed as temporal averages of a series of events. To localize the centroid of the person, firstly silhouette is enclosed by minimum bounding rectangle and then the centroid point (x_c, y_c) of the bounding box is marked in each frame:

$$x_c = \left[\sum_{i=1}^{N} x_i \Big/ N \right] \qquad y_c = \left[\sum_{i=1}^{N} y_i \Big/ N \right] \tag{6}$$

The sequence of centroid points over successive frames forms this part of our feature vector. The curves obtained from sequence of successive centroid positions are shown in Fig. 5. Notice that the movement of centroid is computed for a 2 second duration. We can infer from Fig. 4 that the trajectory of centroid is completely different in ten scenarios. For example when fall starts, a major change occurs in vertical direction and the centroid point has meaningful fluctuations, but while walking or running movement of centroid is not so sensible.

C. Posture Classification

The Support Vector Machine (SVM) is a new and promising classification technique. Basically, there are two types of approaches for multi-class SVM. One is considering all data in one optimization formulation; the other is constructing and combining a series of binary SVMs, such as One-Against-All (OVA), One-Against-One (OVO) and DAGSVM. OVO is constructed by training binary SVMs between pairwise classes. Thus, OVO model consists of $k(k-1)/2$ binary SVMs for K-class problem. Each of the $k(k-1)/2$ SVMs casts one vote for its favored class, and finally the class with most votes wins. Furthermore, to resolve unclassifiable regions and improvement of recognition rate, we introduce fuzzy version of OVO SVM.

1) Conventional OVO SVM

Although pairwise SVMs reduce the unclassifiable regions but they still exist [16]. In pairwise SVMs, we determine the decision functions for all the combinations of class pairs. In determining a decision function for a class pair, we use the training data for the corresponding two classes. Let the decision function for class i against class j, with the maximum margin, be:

$$D_{ij}(x) = W_{ij}^T(x) + b_{ij} \tag{7}$$

Where W_{ij} is the l- dimensional vector, $g(x)$ is a mapping function that maps x into the l- dimensional feature space, b_{ij} is the bias term, and $D_{ij}(x) = -D_{ij}(x)$. The regions $R_i = \{x \mid D_{ij}(x) > 0, j = 1,2,...,n, j \neq i\}$ do not overlap, and if x is in R_i, we classify x into class i. If x is not in R_i , x is classified by voting. Namely, for the input vector x, $D(x)$ is calculated as follows:

$$D_i(x) = \sum_{j=1, j \neq i}^{n} Sign(D_{ij}(x)) \tag{8}$$

And x is classified into class:

$$\arg \max_{i=1,2,..n} D_i(x) \tag{9}$$

If $x \in R_i$, $D_i(x) = n-1$, $D_k(x) < n-1$ $for\, k \neq i$, then x is classified into i. But if any of $D_i(x)$ is not $n-1$, equation (9) may be satisfied for multiple i s, so x is unclassifiable.

2) Fuzzy OVO SVM

To resolve unclassifiable regions we introduce the membership function. To this aim, for i^{th} optimal separating hyperplane $D_i(x)=0\ (i \neq j)$ one-dimensional membership functions $m_{ij}(x)$ are defined in the directions orthogonal to $D_i(x)=0$ as follows:

$$m_{ij}(x)=\begin{cases}1 & if\ D_{ij}(x)\geq 1 \\ D_{ij}(x) & Otheerwis\end{cases} \tag{10}$$

Class i of membership function of x is defined by the minimum or average operation for $m_{ij}(x)\ (j\neq i, j=1,2,\ldots n)$

$$m_i(x)=\min_{\substack{j=1,2,..n \\ j\neq i}} m_{ij}(x) \tag{11}$$

$$m_i(x)=\frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq i}}^{n} m_{ij}(x) \tag{12}$$

Now an unknown datum x is classified into the class:

$$\arg\max_{i=1,2,..n} m_i(x) \tag{13}$$

Equation (11) is equivalent to

$$m_i(x)=\min(1, \min_{\substack{j=1,2,\ldots n \\ j\neq i}} D_{ij}(x)) \tag{14}$$

Because $m_i(x)=1$ holds for only one class, classification with the minimum operator is equivalent to classifying into the class:

$$\arg\max_{i=1,2,..n}\ \min_{\substack{j=1,2,\ldots n \\ j\neq i}} D_{ij}(x) \tag{15}$$

So the unclassifiable region is resolved for the fuzzy SVM with the minimum operator.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Data Acquisition

In order to validate the overall system performance we applied the proposed approach to a set of videos recorded in our lab. The dataset has been collected along two weeks, by considering different light and weather conditions. 48 subjects with different height, weight and genders whose ages ranged from 20 to 30 were asked to participate in the project. We repeated 10 kinds of activities by 5 times in the experimental space and finally 50 video clips were captured and recorded in AVI format with a resolution of 320*240 pixels at 30 frames per second. Figure 4 illustrates examples of each motion. Our dataset for our experiment is composed of video sequences representing three main categories:

a) **Normal Daily Activity:** Five different normal daily activities have been considered: Walking with normal speed, Running, Bending down for catching something on the floor and then rising up, Sitting down on the floor and then standing up and Lying down on the floor.

b) **Unusual Behavior (Fall):** Although the scenarios of falling are very various, we can categorize them in three classes. Detection of various types of falls is valuable in clinical studies to determine fall incidence and costs associated with the treatment of fall injuries. As most falls occur during intentional movements initiated by the person, they happen mainly in forward or backward: stumbling on an obstacle during walking, backward slip on wet ground. But in some cases the fall occurs sideways, either during a badly controlled sit to stand transfer. In this case, the person frequently tries to grip the wall [1].

c) **Abnormal Gait:** Subjects were asked to walk in two special unusual ways: **Stumbling** and **Limping**. In the first case subjects were asked to walk in an unusual way, e.g. as if they were suffering a balance deficiency such as dizziness. In these situations person is unable to keep his balance or symmetry and synchrony of his movement. Body movements suggest the person is in a dubious condition. Whereas the case of limping may be caused by unequal leg lengths, experiencing pain when walking, muscle weakness, disorders of proprioception, or stiffness of joints. Someone taking a step with a limp appears to begin to kneel and then quickly rise up on the other leg; to bystanders the process may appear arduous and painful.

B. Performance Evaluation

The experimental results show that the system has a robust recognition rate in detecting occurrence of considered events. Table I represents the experimental results. N_a refers to number of actions. N_c is number of correctly detected events, N_f is number of falsely detected events and R is the recognition rate.

TABLE I
RECOGNITION RATE OF DIFFERENT BEHAVIORS

Events	N_a	N_c	N_f	R
Walk	120	108	12	90.00
Run	120	104	16	86.66
Stumble	120	100	20	83.33
Limp	120	101	19	84.16
Forward Fall	120	109	11	90.83
Backward Fall	120	112	8	93.33
Sideway Fall	120	104	16	86.66
Bend down	120	102	18	85.00
Sit down	120	103	17	85.83
Lie down	120	114	6	95.00

Also, for better understanding wrong classification results, confusion matrix of classifier output is illustrated in Table II. Notice that E1-E10 are respectively representative of these events: Walk, Run, Stumble, Limp, Forward Fall, Backward Fall, Sideway Fall, Bend down, Sit down and Lie down.

TABLE 2
CONFUSION MATRIX

	Classified as									
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
E1	108	6	4	2	0	0	0	0	0	0
E2	9	104	4	3	0	0	0	0	0	0
E3	10	6	100	4	0	0	0	0	0	0
E4	8	5	6	101	0	0	0	0	0	0
E5	0	0	0	0	109	0	6	4	1	0
E6	0	0	0	0	0	112	4	0	0	4
E7	0	0	0	0	8	4	104	4	0	0
E8	0	0	0	0	2	4	6	102	6	0
E9	0	0	0	0	5	0	4	6	103	2
E10	0	0	0	0	0	4	2	0	0	114

V. CONCLUSIONS AND FUTURE WORK

In this paper a novel efficient approach for activity recognition, principally dedicated to fall detection is proposed. The combination of motion and change in the human shape gives crucial information on human activities. Our experiments indicate that fuzzy multi-class SVM method is more suitable for human motion recognition than the other methods. Our fall detection system has proven its robustness on realistic image sequences of ten different normal, abnormal and unusual human movement patterns. Reliable average recognition rate of experimental results (88.08%) underlines satisfactory performance and efficiency of our system. Moreover while existing fall detection systems are only able to detect occurrence of fall behavior, the proposed system is able to detect type of fall incident (forward, backward or sideway).

Future works will include the incorporation of multiple elderly monitoring which is able to monitor more than one person in the scene and also be able to handle occlusion. Using additional features is also a subject to be explored in the future work. Although the human shape has thus far proven to be a useful method for feature extraction; additional features may be necessary or useful.

REFERENCES

[1] N Noury, A Fleury, P Rumeau, A Bourke, G Laighin, V Rialle, J Lundy "Fall detection - Principles and Methods" Conf Proc IEEE Eng Med Biol Soc. 2007 ;1 :1663-1666 18002293 (P,S,E,B)

[2] Arie Hans Nasution and S. Emmanuel; Intelligent Video Surveillance for Monitoring Elderly in Home Environments, International Workshop on Multimedia Signal Processing (MMSP), Greece, October 2007.

[3] Sixsmith A, Johnson N. "Smart sensor to detect the falls of the elderly", IEEE Pervasive Computing, vol. 3, no. 2, pp. 42–47, April-June 2004.

[4] Caroline Rougier, Jean Meunier, Alain St-Arnaud, Jacqueline Rousseau, "Fall Detection from Human Shape and Motion History Using Video Surveillance" Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference on, Vol. 2 (2007), pp. 875-880.

[5] B. T¨oreyin, Y. Dedeoglu, and A. C¸ etin. HMM based falling person detection using both audio and video. In IEEE International Workshop on Human-Computer Interaction, Beijing, China, 2005.

[6] S.-G.Miaou, P.-H. Sung, and C.-Y. Huang, "A Customized Human Fall Detection System Using Omni-Camera Images and Personal Information" Proc of Distributed Diagnosis and Home Healthcare(D2H2) Conference, 2006

[7] R. Cucchiara, A.Pratti, and R.Vezani, "An Intelligent Surveillance System for Dangerous Situation Detection in home Environments" , in Intelligenza artificable, vol.1, n.1, pp. 11-15, 2004

[8] T. Lee and A. Mihailidis. An intelligent emergency response system: preliminary development and testing of automated fall detection. Journal of telemedicine and telecare, 11(4):194–198, 2005.

[9] Nait CH, McKenna SJ. "Activity summarisation and fall detection in a supportive home environment", Int. Conf. on Pattern Recognition (ICPR), 2004.

[10] Tao, J., M.Turjo, M.-F. Wong, M.Wang and Y.-P Tan, "Fall Incidents Detection for Intelligent Video Surveillance" , ICICS2005

[11] C.Rougier, J.Meunier, A. St-Arnaud, J.Rousseau, "Monocular 3D Head Tracking to Detect Falls of Elderly People", International Conference of IEEE Engineering in Medicine and Biology Society, Sept 2006

[12] Lin,C.-W.,et al.,"Compressed-Domain Fall Incident Detection for Intelligent Home Surveillance". Proceedings of IEEE International Symposium on Circuits and Systems,ISCAS 2005,2005:p.2781-3784.

[13] L. Wang, From blob metrics to posture classification to activity profiling, Proceedings of the 18th International Conference on Pattern Recognition (ICPR, Poster), IV: 736-739, 2006

[14] A. Madabhushi and J.K. Aggarwal, "A Bayesian approach to human activity recognition," Proc. Second IEEE Workshop on Visual Surveillance, pp.25–32, June 1999.

[15] M.Shakeri, H.Deldari, H.Foroughi, "A Novel Fuzzy Background Subtraction Method Based on Cellular Automata for Urban Traffic Applications", 9th International IEEE Conference on Signal Processing, ICSP'08

[16] S.Abe, S.Singh, Support Vector Machine for Pattern Classification, Springer-Verlag London Limited 2005

Limits in Process Diagnosis Using Thermographic Images Processing

Alina Dinca[1], Constantin Cercel[2], Ilie Borcosi[3]
"Constantin Brancusi" University
Geneva street, no.3, Targu-Jiu, Romania
[1]alina@utgjiu.ro, [2]costi@utgjiu.ro, [3]ilie_b@utgjiu.ro

Abstract – **Thermography is usually used to observe processes' evolution, to make predictions about a process, to diagnose a process and so on. In this paper the authors present a simple application which processes data from thermographic images automatically in order to draw conclusions about the observed processes (to diagnose the process). The application is tested on thermographic images taken from different processes and the accent is put on the limits imposed by each process in creating a general algorithm to diagnose correctly any kind of process. The purpose of this application is to diagnose a process automatically, with no human analysis involved.**

I. Introduction

In order to keep a process within a range of parameters, called nominal parameters, a process observer (human or integrated software) has to keep track of the parameter evolution in time. This is usually done with data acquisition techniques and equipments. With the acquisitioned data about the observed parameters, the observer can establish a diagnosis for the process. So, diagnosis is of the essence for maintaining a process to function in normal conditions. From a diagnosis report a specialist could predict what would happen to certain equipments if no change is made in the process or if changes are made, thus preventing dangerous situations. In big industrial complexes, there are trained specialists who develop their own software for specific applications, function of the processes that needs to be controlled, monitored or diagnosed.

State of the art techniques and methods used for diagnosis depend on the observed process. We could use frequency analysis to determine the technical state of the aggregates, simple parameter measurement to determine nonmeasurable values, thermography as a noninvasive technique where conditions impose or when we have a process in which heat is emitted etc. For example, in the abstract of Reference [1] there are three diagnostic techniques for renewal of electrical equipment installed in public plat such as water purification plant, sewage treatment plant, and waste matter disposal plant. 1. Physical Diagnosis: Visual Check and Measurement with Instruments, 2. Functional Diagnosis: Hearing from Operators of the Plant, 3. Economical Diagnosis: To investigate cost for Construction, Maintenance, and Repair respectively and also total.

In this paper the authors chose to apply and present automated thermographic image processing as a method and for non-destructive detection: heat variations detection in building insulations and maximum heat radiation value detection in electrical equipments for two cases using an infrared camera to capture images of samples that have been studied during process functioning. The authors have taken into account for comparison how would the processes in nonworking state look through the "eyes" of automated thermographic analysis.

Data contained in the thermographic images (called thermograms from this point forward) are very important for diagnosis of processes, predictive maintenance, interfering in the process before equipments or systems fails or stops working. The extracted data can be analyzed with different algorithms, compared, stored in a database, processed, and also can be used to create maps or graphics which present an evolution in time of certain observed parameters, all these without modifying the thermogram. The diagnosis is given by the way these data are interpreted. From the experience the authors have in the IR field, these data are not enough for a programmer to create software which will diagnose correctly any kind of processes.

Every producer and distributor of infrared cameras offers software with the camera at a very high price. From Reference [2] is known that this software is actually a reporter, and it usually extracts and processes all the data found in thermograms for us, and presents them in a technical report for the thermographic inspection. This is about all that an infrared camera reporter can do, so it limits the possibilities of data processing. But a thermographer may need for certain processes (like our case of automated processes, in surveillance or predictive maintenance) more than that. He would need certain data from the thermogram plus additional data which is not given (such as: emissivity for each material in the thermogram, whether there are other sources of thermal radiation around the observed process which do not appear in the thermogram, but have the capacity to influence the interpretation etc) for further processing and correct interpretation of the thermogram.

So the limits come from the fact that each process is characterized by a set of parameters which differ from a process to another. A personal opinion would be that this is the reason the infrared camera distributors only offer a reporter as software for the camera, remaining that each thermographer to use the given data from the thermogram in a particular way for each process in order to draw conclusions and put a diagnose.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_96, © Springer Science+Business Media B.V. 2010

In Reference [3] the author states that the special theme of industrial diagnosis, planning and simulation is structured around a selection of industrial problems rather than a particular type of technology or methodology, so a variety of techniques and applications that are of relevance to these problems must be chosen. The authors would add that the diagnosis is made in order to prevent a problem from appearing in a process, and any problem can be described through variation of certain process parameters outside the normal limits.

II. Data in thermograms

A. *What is a thermogram*

From Reference [6] is known that Infrared Thermography, thermal imaging, or thermal video, is a type of infrared imaging science. Thermographic cameras detect radiation in the infrared range of the electromagnetic spectrum (roughly 900–14,000 nanometers or 0.9–14 µm) and produce images of that radiation. Since infrared radiation is emitted by all objects based on their temperatures, according to the black body radiation law, thermography makes it possible to "see" one's environment with or without visible illumination. The amount of radiation emitted by an object increases with temperature, therefore thermography allows one to see variations in temperature (hence the name). When viewed by thermographic camera, warm objects stand out well against cooler backgrounds; humans and other warm-blooded animals become easily visible against the environment, day or night.

It is important to note that thermal imaging displays the amount of infrared energy emitted, transmitted, and reflected by an object. Because of this, it is quite difficult to get an accurate temperature of an object using this method.

So, a thermoram is a thermal radiation distribution shown in a variety of colors, according to the chosen color palette, not a map showing temperatures. The temperature is usually calculated and displayed by the camera after some internal calculus which takes into account emissivity, reflected and ambient temperature.

B. *The format of thermograms*

Infrared cameras save the thermograms in different file formats, depending on the camera producer, but mainly in .jpeg formats, because this format allows a very good compression without data loss and this because more storage space is available on the camera's memory card.

For example, at page 13 in Reference [4] is noted that the ThermaCAM™ QuickView software (from Flir Systems) supports a variety of different radiometric and non-radiometric file formats. These include not only FLIR Systems propriety file formats, but also a number of third-party file formats.

ThermaCAM™ QuickView supports the following radiometric file formats:

- ThermaCAM radiometric *.jpg
- ThermaCAM radiometric *.img
- ThermaCAM radiometric 8-bit *.tif
- ThermaCAM radiometric 8/12-bit *.tif
- ThermaCAM radiometric 12-bit *.tif
- ThermoTeknix *.tgw
- ThermoTeknix *.tmw
- ThermoTeknix *.tlw

ThermaCAM™ QuickView supports the following non-radiometric file formats:

- *.jpg – as defined by *Joint Photographic Experts Group*. A widely used image format that is compatible with the majority of all image software on the market. It takes advantage of a powerful compression algorithm named after the committee that defined it.
- *.bmp – as defined by Microsoft®. This is one of the most wide-spread file formats for images and is supported by the majority of image browsers and wordprocessing programs. However, the *.bmp file format creates considerably larger files than the *.jpg file format.

As in Reference [7], arrangement of detector and image data of infrared thermography is shown in the figures below:

Temperature distribution image data (fig. 2):

Temperature distribution image data of infrared thermography consists of matrix of pixels (number of detector: for example, 320 horizontal x 240 vertical pixels) as shown in fig. 2. Thermal image data can be transferred to PC. Subsequently, the data can be calculated and utilized freely. Thermal image data is colored up pixel by pixel based on temperature, in fact on thermal radiation distribution.

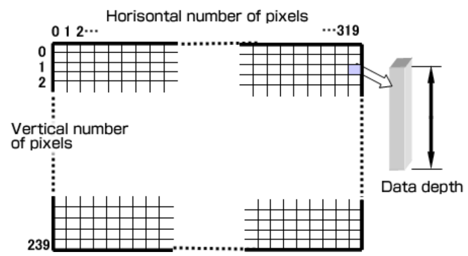


Fig. 1

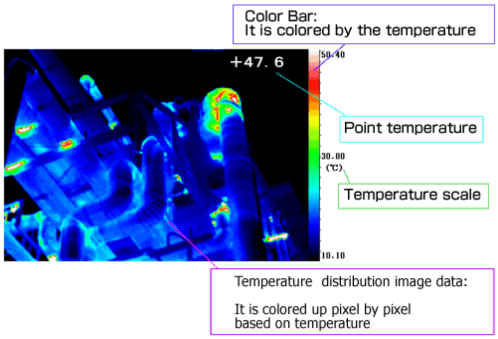


Fig. 2

A thermogram contains as any .jpeg, the information in the image in pixels, and also some metadata about the camera model, photographer, resolution etc., function of the type of camera and what it can put as metadata.

Some of the metadata extracted with specific algorithms and can be used for image processing, but most of the necessary information for image processing is found within the pixels of the thermogram.

C. *Transfer of thermographic images*

Connecting the camera to a PC is done as in fig. 3, as presented in Reference [4].

The camera can be connected to a laptop or desktop computer using one of the following three communication protocols:
- RS-232 serial communication protocol
- USB Universal Serial Bus communication protocol
- FireWire communication protocol

Different camera support different communication protocols. For example, a ThermaCAM S60 will support both USB and FireWire, while a ThermaCAM E2 will support USB and RS-232.

In fig.3. the callouts are:
1 – FireWire communication cable.
2 – RS-232 communication cable
3 – USB communication table
4 – USB communication table
5 – RS-232 communication cable

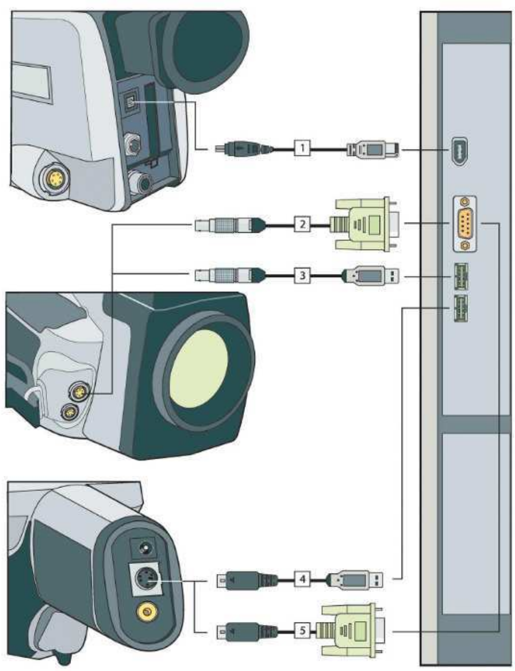


Fig. 3

III. Application tested on several thermograms

For our application we used the principle of measurement by Infrared Thermography equipment found in Reference [6], and represented in fig. 4.

In reference [4], on pages 79-80 there are some Thermographic measurement techniques explained in the following rows.

An infrared camera measures and images the emitted infrared radiation from an object. The fact that radiation is a function of object surface temperature makes it possible for the camera to calculate and display this temperature. However, the radiation measured by the camera does not only depend on the temperature of the object but is also a function of the emissivity. Radiation also originates from the surroundings and is reflected in the object. The radiation from the object and the reflected radiation will also be influenced by the absorption of the atmosphere. To measure temperature accurately, it is therefore necessary to compensate for the effects of a number of different radiation sources. This is done on-line automatically by the camera. The following object parameters must, however, be supplied for the camera:
- The emissivity of the object
- The reflected temperature
- The distance between the object and the camera
- The relative humidity

The most important object parameter to set correctly is the emissivity which, in short, is a measure of how much radiation is emitted from the object, compared to that from a perfect blackbody. Normally, object materials and surface treatments exhibit emissivity ranging from approximately 0.1 to 0.95. A highly polished (mirror) surface falls below 0.1, while an oxidized or painted surface has much higher emissivity. Oil-based paint, regardless of color in the visible spectrum, has an emissivity over 0.9 in the infrared. Human skin exhibits an emissivity close to 1. Non-oxidized metals represent an extreme case of almost perfect opacity and high spectral reflexivity, which does not vary greatly with wavelength. Consequently, the emissivity of metals is low – only increasing with temperature. For non-metals, emissivity tends to be high, and decreases with temperature.

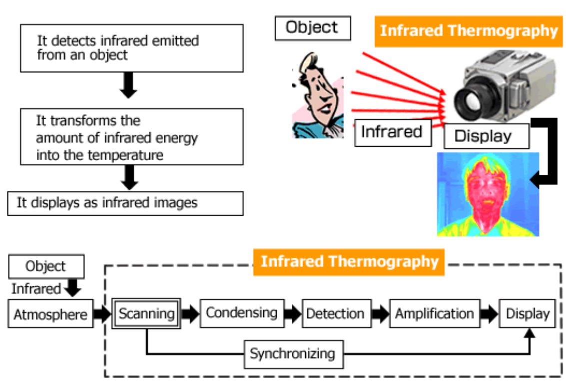


Fig. 4

As the first author learned from [5], the following applications were studied in order to observe and diagnose two processes:

- cold air infiltration in walls or heat loss;
- a high voltage switch during functioning who's upper region was heated on purpose in comparison with a panel with different materials.

A. The software used for the applications

We used LabWindows CVI as software to implement the algorithm which decides whether there is a problem and where It works in the following way:

- first, a thermogram is loaded.
- a threshold is set
- after that, the thermogram is scanned at a bit level with the purpose of finding color maximums (maximum of Red, of Green and Blue) above threshold.
- with the obtained maximums we set a new image only in black and white.
- the obtained image is displayed in the right side of the panel.
- in the obtained image will only be distinguished the areas which are of the interest for the observer (the areas with air infiltration problems, respectively the areas with maximum values of thermal radiation, which theoretically means that they have the highest temperature if the emissivity is set correctly)

All these operations happen after pressing the "OK" button on the panel.

B. Application for thermal radiation variation

Thermograpy as diagnosis technique is used more and more for checking building insulation, infiltrations, or heat loss in different areas of the constructions domain.

Analyzing a thermogram showing a house interior wall, we need to find out without looking at the picture if there is heat loss or if there are cold air infiltrations into the room and the position of that area in the thermogram. In order to develop a simple program which would determine the "if" and "where", the algorithm should only decide if there is a difference of thermal radiation in the thermogram and to highlight that area, and the algorithm described at subsection A fits the description.

On the left of fig. 5. the thermogram of the wall can be seen, and in the right side is the resulted image after being processed with the algorithm described above.

As seen in the left image of fig. 5., the areas drawn with a lighted color represent the thermal radiation of the wall. If the color is the same all over the observed wall (thermogram), the thermographer or software puts the following diagnosis: the insulation of the wall is good, and there are no air infiltrations. This is not our case. In our case, in the middle of the thermogram, upper-right corner and bottom-right corner, the human eye detects a sudden variation of color, which leads to the conclusion that in that area there is a problem (cold air infiltration, or heat loss, because it appears in the range of colors closer to darker colors, which mean by default, colder spots). Also, applying the algorithm, in the right picture we can see in the locations mentioned above (middle, upper-right corner, bottom-right corner) that a zone of white appears, signaling in that exact position that there is a problem. A thermograms without problems would be all black.

C. Application for finding the maximum temperature value

In electrical processes we use thermography because of the advantages it offers compared to other investigation techniques. In this paper, for the second application, I chose a lab platform used in [5] which shows a high voltage switch that was heated on the upper side on purpose, in order for us to interpret the thermogram and diagnose the process it is used in.

Where the value of the color in the thermogram exceeds a certain threshold, the area containing it will be reproduced in the result picture. The threshold must be given function of the nominal parameters at which the equipment works.

In this case, using the same algorithm for the first application, the authors observe the following:

- the result given by the algorithm is wrong, because it shows that there are problems not only where they really are, but also in other areas. In this case, the algorithm we designed is limited and must be improved by making the threshold variable and to be introduced by the operator before thermogram processing.
- because of the reflected radiation from other warm or hot bodies, we can see on the thermogram a close to color pink which doesn't have the temperature corresponding to its color. The specialist knows that. The algorithm doesn't, thus shows the zone in the result picture as being a problem.

Applying the algorithm to thermogram in fig. 7, we will obtain a worse result than in the previous application. This is because we applied an algorithm which indicates maximum values on a thermogram whose role is to point out that for different materials which are heated at the same temperature, the temperature scale on the thermogram will vary, though it should look the same. And this is another limit of our algorithm: for each thermogram the software needs to know the emissivity of each material situated in the thermogram, which is difficult to find out even by a thermographer. Even he would need to take several thermograms and to calculate several times in order to know the emissivity for each material.

Also, the distance must be taken into account, because the bigger the distance of measurement, the higher the risk to make wrong interpretations of the thermograms, thus resulting a wrong diagnosis.

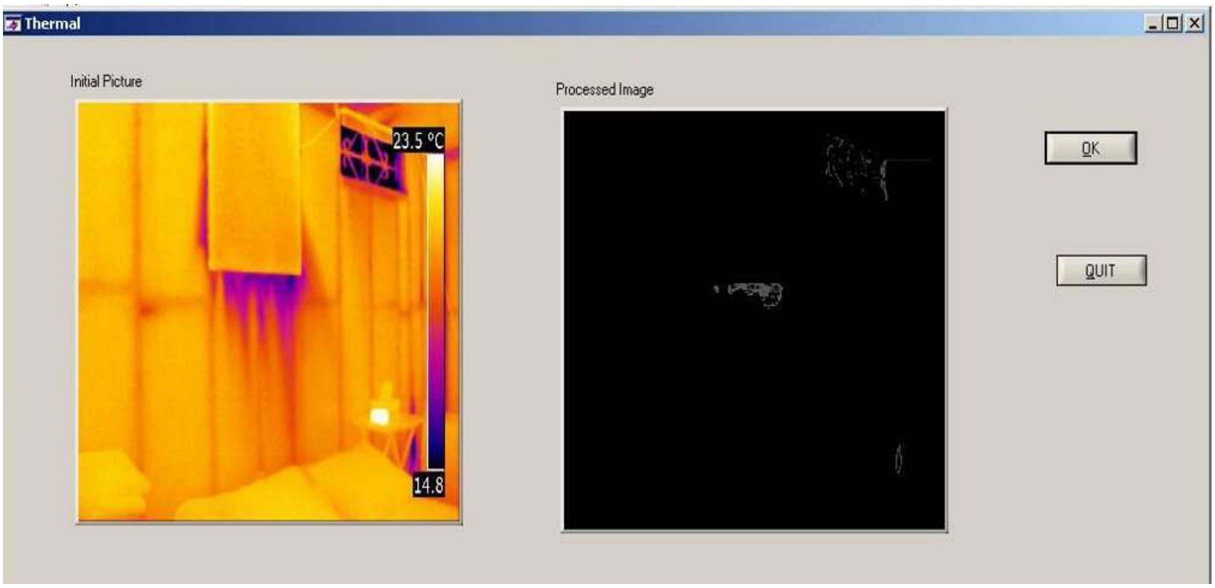


Fig. 5. Application for room wall thermogram and result of its processing

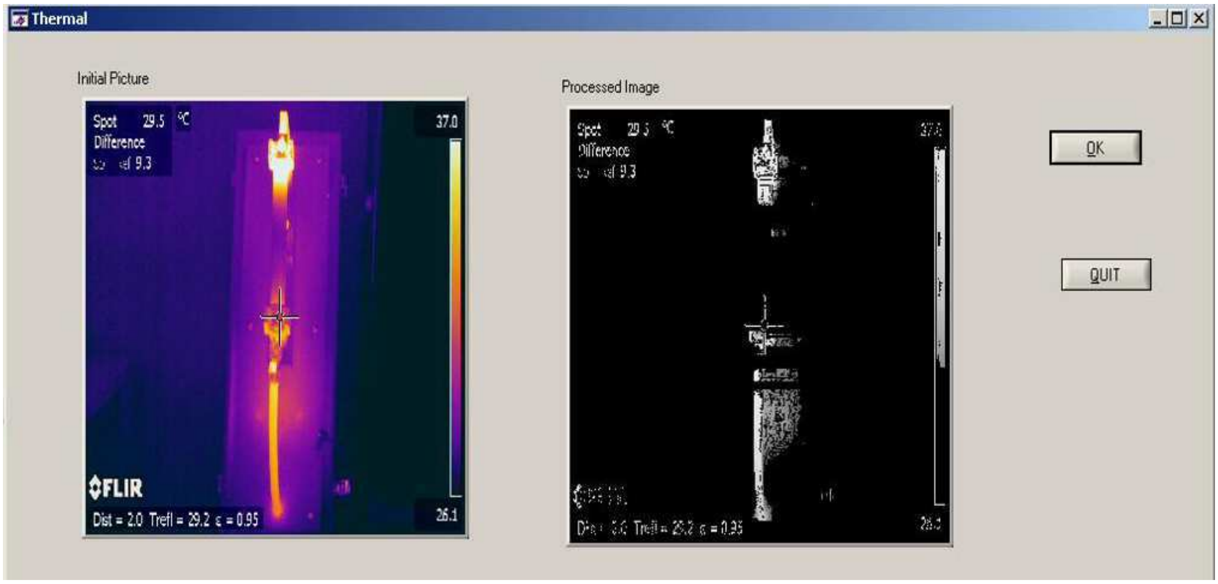


Fig. 6. Application for high voltage switch thermogram and result of its processing[5]

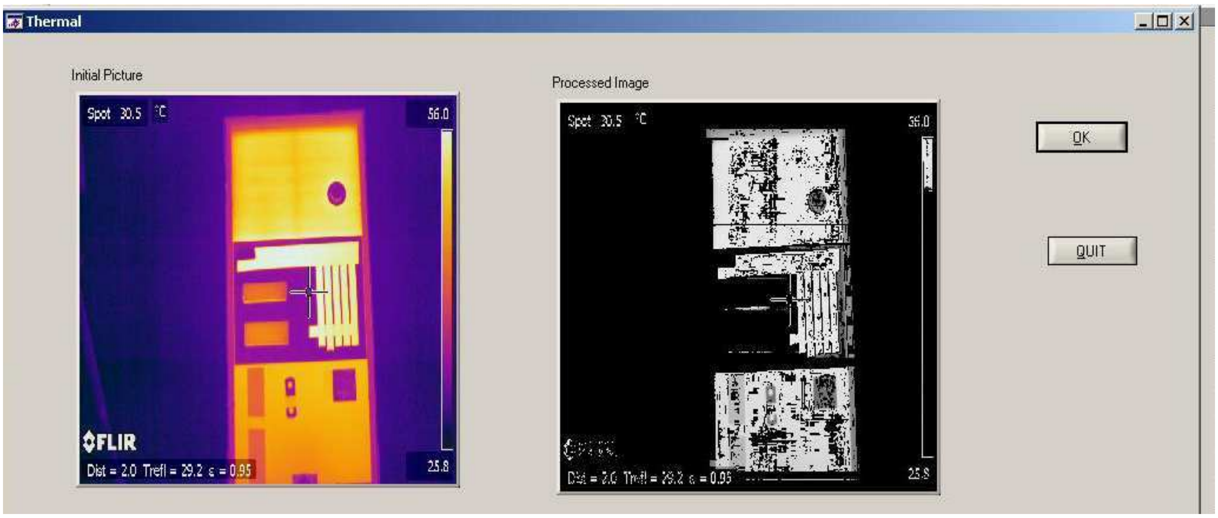


Fig. 7. Application for a panel with different materials thermogram and result of its processing [5]

IV. CONCLUSIONS

In this paper we only pointed out few of the limits met in algorithms for thermographic image processing in process automated diagnosis. The author concluded from the results they obtained that either a more complex algorithm should be designed containing as many extra data included before taking the picture in order to be used for as many applications as possible, or the algorithms we designed for each process have the disadvantage that may not be used for other process' diagnosis.

The algorithms are simple and still need more adjustments.

Also, from the authors' point of view, a thermography expert's eye joined with his experience give a much better interpretation, thus a better diagnosis than an automated one (with computer algorithm). The thermographer must assure that he sets all the parameters for the camera to give a very suggestive thermogram.

REFERENCES

[1] Shimaoka Masahiro (Mitsubishi Electric Corp., JPN) and Takakura Toshikazu (Mitsubishi Electric Corp., JPN) – "Examples of the electric equipment diagnosis techniques for water purification plant, sewage treatment plant and waste matter disposal plant," *Papers of Technical Meeting on Public Plant Engineering, IEE Japan,* Code: L0939A, Vol. PPE-02, No. 9-15, Page 25-28, 2002.

[2] Flir Systems – "ThermaCAM Reporter 7 User's Manual".

[3] Kreuger, Per – "Industrial diagnosis, planning and simulation: introduction and key note to special theme," *ERCIM News-online edition* (no. 56). Pp 10-11, January 2004.

[4] Flir Systems - User's Manual for ThermaCam^{TM} Quick View, version 1.1+

[5] ITCC Training Course for Thermographer level I, Bucure□ti, Romania (Pictures taken with Flir Infrared cameras B and T series, property of ITCC Romania, during the Laboratory of the Infrared Training Course, level I thermographer.).

[6] http://www.nec-avio.co.jp/en/products/ir-thermo/what-thermo.html

[7] http://en.wikipedia.org/wiki/thermography

E-learning Tools in Naval Engineering

G. Popescu,
Naval Architecture Faculty
„Dunarea de Jos" University
Galati, ROMANIA
gabriel.popescu@ugal.ro

Abstract- **The designing tools for an e-learning system in the naval engineering are the argument of this paper. The field particularity, like variety and size of ship structures and piping have defined the methods and techniques for developing the specific components of the e-learning system, based mainly on the generation of animated virtual models. Therefore the creation of the necessary tools for a correct 3D perception of the ship and also the tools for self assessment and assessment are the main components for the present e-learning field. The efficient use of the e-learning tools is provided by the management of a unitary platform as an integrated e-learning system.**

Testing and validation of the e-learning tools has been an important phase in implementing the educational system in the Naval Architecture Faculty of Galati.

Keywords: ***e-learning tools, naval engineering, typical problems, simulation, testing, validation***.

I. INTRODUCTION

In engineering education, the curriculum requires specific technical subjects. The time budget is mostly assigned for laboratories and projects.

If in the electronic and electrical engineering or computer science, teaching materials are basically circuits and components, in the naval education system, the ships, blocks of ships and their panels can be presented to the students only in shipyards. Because it is impossible to carry out all laboratories and project hours in the shipyard, we had to realize the photos and layout models of the ship's structures and piping.

Such materials have subsequently been replaced by electronic documents like .doc, .pdf and .ppt files. This list has been completed by video files with the recordings of welding and assembly processes.

All these materials have been used sequentially in the educational process.

Another feature in preparing naval engineering students is related to the ship manoeuverability study. If ship structures can be presented in shipyards, manoeuverability has been presented only in theory because teaching on board at sea is impossible.

As regards assessment, students were traditionally evaluated by classical examination forms with three subjects, two argumentative and one problem.

II. PROBLEM FORMULATION

The quality of the education system is our main concern therefore we consider it is very important to design, execute and implement an integrated e-learning system.

E-learning designing platform has to meet the exigencies of this particular field of education and the student's training needs, as well as quality teaching standards.

A successful e-learning system is based on the formulation of certain technical specifications.

The system must:

have the characteristics and properties of an integrated system

- contain tools for presenting text and hypertext
- contain presentation tools for 2D and 3D models
- contain presentation tools for video files
- contain self - assessment tools
- contain assessment tools
- provide facilities for centralizing the assessment and statistic presentation.

Fig. 1 The menu bars

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_97, © Springer Science+Business Media B.V. 2010

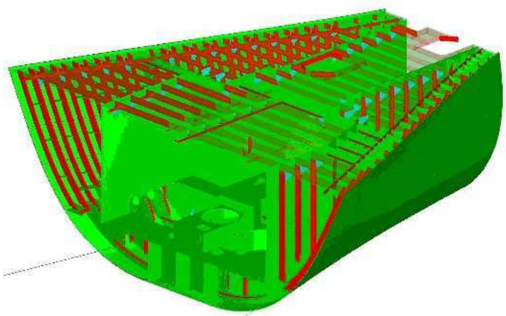


Fig.2. Virtual model of a block ship structure

- respect psychological and pedagogical recommendations on visual perception and assimilation
- provide free or guided training
- allow teacher intervention in the system to update the information
- have interfaces with shape patterns, chromatic and dynamic, to be consistent with trends in web-art-design.

III PROBLEM SOLUTION

For application development purposes, we chose Visual Basic platform, due to it being situated in the top generators of the layout's manager. E-learning system has been designed like any classical Windows application and as a result students are very familiar with it.

Such a system contains the first headline with the name of the discipline, followed by a line for menu bars, designed for discipline chapters. Each menu opens submenus particular to discipline subchapters, which in turn contain submenus with various information type arguments, examples, simulations, tests and self-assessment tests (see Fig.1**)**.

If for the shipbuilding discipline, the e-learning system needs tools for virtual models of various types of ship structures (see Fig.2) that students can "handle" for the maneuverability course, the tools for 2D and 3D simulations are required, as well.

In the case of the turning circle maneuver, the e-learning system was designed to open two windows simultaneously (see Fig.3). In the lower window the maneuver of 2D turning is presented, and in the upper window students can see the virtual reality "being" on the command post of the ship. The 2D simulation is the result of the transposition of turning circle equations of the ship in AutoLISP code file, creating and representing the instant ship positions.

Virtual reality has been conceived by 3D modeling of the ship and the port area, in the 3DStudioMAX system, defining the ship's movement curves produced in an animated system. Students can see the turning circle maneuver as they would be on board, simultaneously visualizing in the 2D and 3D windows which form a virtual reality.

Self-assessment tests were created in Java using items form. The instantaneous report (see Fig. 4), which displays if

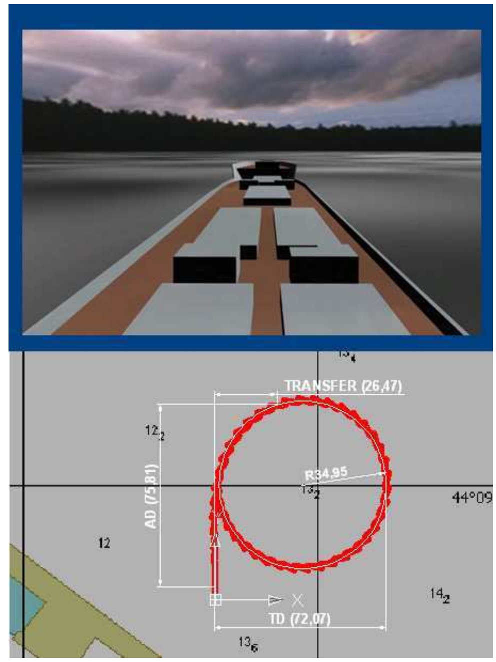


Fig.3. E-learning tools for 2D and 3D visualization

the response was correct or incorrect, is very beneficial to the student.

Self-assessment tests can be found in the structure of the e-learning system in each chapter and section, thereby allowing students to set the new arguments during the teaching.

Assessment tests have a summative nature and have been generated in Java and are divided into 2 types, items with one correct answer and with multiple correct answers.

The results of assessment tests are viewed primarily by the student. At the end of each test the system informs the student giving the number of points obtained (see Fig.5). The results are sent in the system database, in the location reserved for each student, allowing the student to access the test solutions. The tabulation of the results into the same unique database permits to obtain information about individual achievements of each student (see Fig.6) and for the whole group, thus being a tool for teacher self standing (see Fig.7) .

IV TESTING SYSTEM

Like any product, this e-learning application has undergone several tests.

Operational Experiments

They were focused in two directions, according to teacher's observations and student reaction in front of information. Such types of experiments in which teachers checked:

1. Time

- to browse information guided by the teacher
- fixing the concepts taught
- assessment

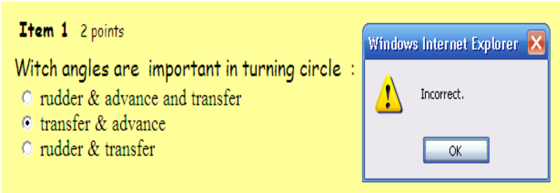


Fig.4. Instant response in self-assessment tests

Fig.5. E-learning system tool for assessment

2. Fluidity

- in presentation the course chapters
- of arguments fixing
- assessment

3. Application efficiency

Experiments were conducted using two subjects groups: 3rd year students in Naval Architecture Faculty of Galati (2007-2008 academic year) (Bologna education plan).

Original conditions were very close:

- Same specialization (the number of hours)
- Same teacher
- Annual average for two groups 7.26, 7.32

Under those initial conditions it was decided to carry out experiments in the next version:

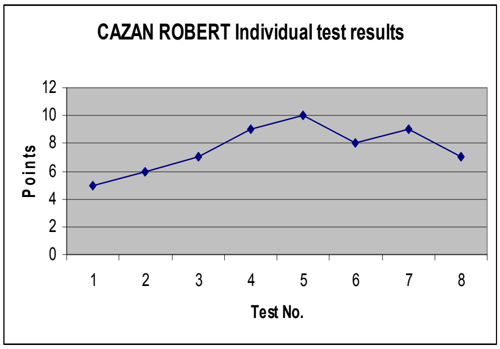


Fig.6. E-learning information facilities for individual achievements

The experiment was applied to the course in maneuverability.

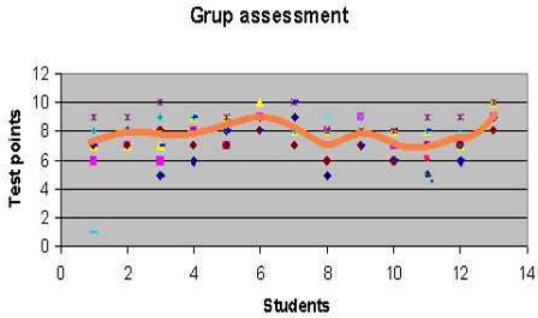


Fig.7. E-learning information facilities for the group achievements

TABLE I
Educational technologies

Group	*Group A*	*Group B*
Ways and Method	Classical	E-learning

If for A group classic educational technologies in teaching and assessment were used, for B group the E-learning system was used.

Results:

• On teaching:

An increased interest for B group versus A group, confirmed by attractive teaching methods (presentation on computer).

Browsing a much larger number of examples (of these issues) for group B vs. A.

Setting concepts more efficiently by the teacher

• Self – assessment :

Fixing the self-taught concepts was possible only for the B group.

• Assessment :

The same assessment tests (items types) were carried out for both groups. They contained questions that have pursued a number of seven standard parameters

1. Issues recognition
2. Forms association, arguments and technologies
3. Accuracy and rigor
4. The capacity of synthesis.
5. Error identification.
6. Formula identification
7. Formula knowledge

TABLE II
Comparative results of correct answers

Comparative parameters Item No.	A group correct answers [%]	B group correct answers[%]
1	72	90
2	68	84
3	65	96
4	82	87
5	66	93
6	79	81
7	83	85

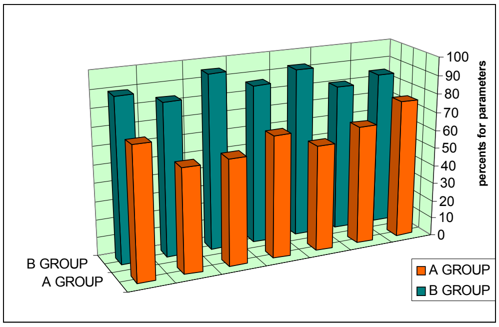


Fig. 8 Graphical comparison between effects of educational technologies

The answers to these questions were the parameters of comparison between the two groups (see fig.8) as follows:

There were important differences in general (in favor of B group) and in particular for the visual perception objectives and imagination.

Explanation for this difference is due to:

• simulation of the ship maneuvers
• dynamic teaching in e-learning system
• e-learning system facilities for student in self assessment
• assessment with exclusion of the subjective factor
•dynamic image viewing and a very large number of examples.

V CONCLUSIONS

There seem to be several reasons for our integrated e-learning system success:

- The interface designed in the context of the learner's tasks, for the purpose of supporting tasks to be done by the learner,
- The interface designed are to latest theories in web-art-design
- Using the visual 2D and 3D perception like a language for attention, understanding and motivation to acquire and properly use knowledge,
- The complex systems of ship structures, piping and maneuverability simulated in virtual reality
- Experience in naval architecture teaching

This e-learning system allows to analyze and to evaluate the performance of lecturers in their classroom thus improving their teaching skills and contributing to their professional training.

Results of the experiment show statistical significance in favor of implementing e-learning system in our faculty, in terms of usability parameters; effectiveness, efficiency, and subject satisfaction.

REFERENCES

[1] Beard, C. (2007) *Experiential learning: The Development of a Framework for Effective Practice*. Unpublished thesis, (PhD). Sheffield Hallam University.

[2] Beard, C., Clegg, S. and Smith, K. (2007) Acknowledging the Affective in Higher Education. *British Educational Research Journal,* 33(2), April, 235-252.

[3] Beard, C. and McCarter, R. (2005) *Mastering University*. [CD-ROM]. Aldershot: Gower.

[4] Beard, C. M. and Wilson, J. P. (2002) *The Power of Experiential Learning: A Handbook for Educators and Trainers*. London: Kogan Page.

[5] Beard, C. and Wilson, J. P. (2005) Ingredients for Effective Experiential Learning: the learning combination lock. *In*: P. Hartley *et al.* (eds) *Enhancing Teaching in Higher Education: New Approaches for Improving Student Learning*. York: Routledge/Higher Education Academy, 3-15.

[6] Beard, C. and Wilson, J.P. (2006) *Experiential Learning: A Best Practice Handbook for Educators and Trainers*. London: Kogan Page.

[7] Clark, R. C. and Mayer, R. E. (2003) *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multi-media Learning*. San Francisco CA: Jossey-Bass/Pfeiffer.

[8] Collis, B. and Moonen, J. (2001) *Flexible Learning in a Digital World: Experiences and Expectations*. London: Kogan Page.

[9 *Beard, Wilson and McCarter (2007) Towards a Theory of e-Learning: Experiential e-Learning, Journal of Hospitality, Leisure, Sport and Tourism Education, Vol. 6, No. 2.,www.heacademy.ac.uk/hlst/resources/johlste*

[10] Fenwick, T. J. (2003) *Learning Through Experience: Troubling Orthodoxies and Intersecting Questions*. Malabar, Florida: Krieger Publishing Company.

[11] Hede, A. (2002) An Integrated Model of Multimedia Effects on Learning. *Journal of Educational Multimedia and Hypermedia, 11 (2), 177-191.*

[12] McCarter, R. (2005) *Analysing the pedagogical value of the synergy between video and text in a digital media application.* Unpublished Masters Dissertation, Sheffield Hallam University.

[13] Moreno, R., Mayer, R. E., Spires, H. A. and Lester, J. C. (2001) The Case for Social Agency in Computer-Based Teaching: do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction,* 19(2), 177-213.

[14] Mortiboys, A. (2002) *The Emotionally Intelligent Lecturer*. Birmingham: SEDA Publications.

[15] Nichols, M. (2003) A Theory for ELearning. *Educational Technology and Society,* 6(2), 1-10.

[16] Payne, P. (2002) On the Construction, Deconstruction and Reconstruction of Experience in Critical Outdoor Education. *Australian Journal of Outdoor Education,* 6(2), 4-21.

[17] Ravenscroft, A. (2001) Designing ELearning Interactions in the 21st Century: revisiting and rethinking the role of theory. *European Journal of Education,* 36(2), 133-156.

[18] Robertson, I. (2000) *Mind Sculpture: Your Brain's Untapped Potential*. London: Bantam Books.

[19] Schoenmaker, J. and Stanshev, I. (2000) *Visualisation, Principles and Tools for Instructional Visualisation*. University of Twente: Andersen Consulting-ECC.

[20] Sherry, L. (2003) Issues in Distance Learning. *International Journal of Educational Telecommunications,* 1(4), 337-365.

Augmented Reality Used to Control a Robot System via Internet

R. Tarca, L. Csokmai, T. Vesselenyi, I. Tarca, F. Popentiu Vladicescu
University of Oradea
1 Universitatii st.
410087 Romania

Abstract: **This paper presents the possibilities of using the augmented reality to control a robot system via Internet.**

I. INTRODUCTION

Augmented reality allows the user to see the real world, with virtual objects superimposed upon or mixed with the real world. Therefore, AR supplements reality, rather than completely replacing it. [2]

In most applications, a scene is captured by a camera and additional information is displayed by suitably designed, virtual objects added to the real scene view, giving the user a better perception of the world state.

Azuma [2] defines Augmented Reality (AR) as any system having the following three characteristics:

- Combines real and virtual;
- Is interactive in real time;
- Is registered in three dimensions.

Although AR systems may also augment other human senses, like the auditory or haptic sense, most current systems only implement the visual channel.

An augmented environment for robotics is a set of online tools, databases and managed resources used for remote manipulation of robots, not necessary using VR resources.

According to this statement, the following type of augmented environments can be addressed (by extension from some E-learning terminology presented in [1]):

- Workstation-based robotics - the computer is connected to the robot (or to robots if an adequate interface is available) and the user interacts with robot using the application software [23].
- Network-based robotic systems - the workstation is embedded in a local area network (LAN) and some LAN protocol is used to access the robot from any workstation (the application software being network-based) [23, 16, 18].
- Web-based robotic systems - the communication is based on HTTP protocol [12, 17, 24, 25].
- Internet-based robotic systems - can use any IP-based protocols [3, 10, 20, 23, 28, 31, 32].
- Collaborative environments for robotics - is a groupware paradigm already considered in [5] and [19].
- Distributed environments for robotics - is a paradigm assuming distributed users, resources, and robots. [5, 30].

The following milestones have to be considered in improving the quality of such environments.

Firstly, the network may cause delays and the communication cannot be guaranteed (sometimes an Internet connection is interrupted and a recovery module is necessary).

Secondly, the algorithms implementing compression / un-compression images, collision detection, and send-receive based communication by messages have to be as fast as possible to assure an approximately real-time behavior.

The third difficulty arises when implementing an interface on the client side to support the user in his or her task to manipulate the robot over a remote connection. An additional approach is based on profile user identification and context sensitive local help provided by Java applets, PHP scripts or JavaScript functions.

II. THE TELEROBOT SYSTEM CONCEPT

The concept of "human supervisory control" (Sheridan, 1992) that underlies a telerobot is illustrated in figure 1.

The human operator interacts with the human-interactive computer (HIC). It should provide the human with meaningful and immediate feedback.

The subordinate task-interactive computer (TIC) that accompanies the controlled robot receives commands, translates them into executable command sequences, and controls command execution.

In a supervisory control system the human supervisor has the following functions:

- Planning what task to do and how to do it.
- Teaching the computer what was planned.
- Monitoring the automatic action to make sure all is going as planned and to detect failures.
- Intervening, which means that the operator supplements ongoing automatic control activities, takes over control entirely after the desired goal has been reached satisfactorily, or interrupts the automatic control to teach a new plan.
- Learning from experience to do better in the future.

The role of computers in telerobotics can be classified according to how much task-load is carried compared to what the human operator alone can carry. They can trade or share control. Trading control includes the following cases:

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_98,

- The computer replaces the human. It has full control over the system.
- The computer backs up the human.
- The human backs up the computer.

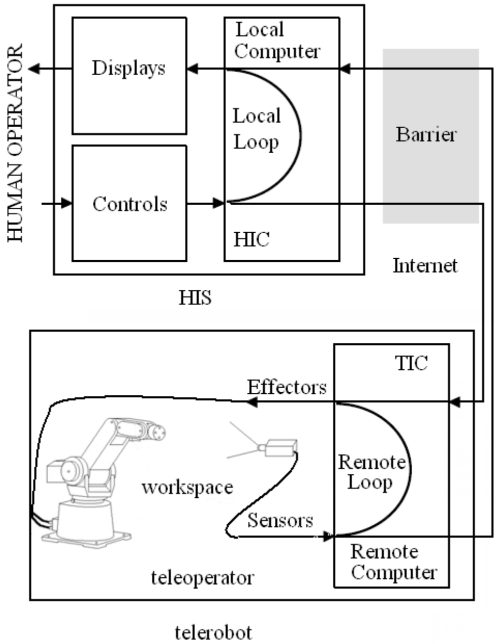


Fig. 1. Basic concept of an Internet telerobot

The most common case in telerobotics is sharing control, meaning that the human and the computer control different aspects of the task:

- The computer relieves the human operator from certain tasks. This is very common in telerobotics when the remote system performs subtasks according to the plans specified by the human operator.
- The computer extends the human's capabilities. This typically occurs in telerobotics when high precision of movements and applied forces is required.

III. The System Structure

The telerobot system developed by our research team is presented in figure 2. The telerobot system structure consists in one server to the remote computer (TIC), and some clients the as local computers (HIC). As a robot we have used a Mitsubishi Movemaster RV-M1 robot with 5 axes.

The structure of the system is similar as the one presented in figure 1.

Different kinds of objects cubes, parallelepipeds are placed on a table, in its workspace. A web camera observes the scene.

The task of the telerobot system is to acquire the scene image with the objects, to calibrate the image, to transfer this image to the client, and then the human operator to generate the 3D model of any piece from the visual field, to overlay this model on the real object, and in this way to obtain the mass centre position and the orientation of the object.

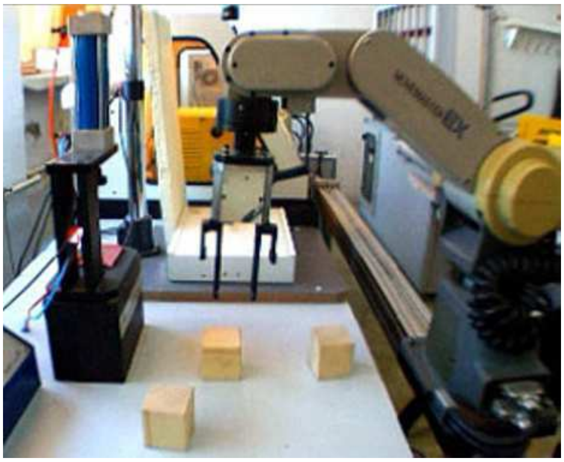


Fig. 2. The telerobot system

With this information, the robot will be driven via Internet to pick the object and place it anywhere in the robot workspace. As it can be seen, different kinds of objects (cubes, parallelepipeds) are placed on the table in the robot workspace.

The images acquired by the web camera are compressed and saved in a file. Each image is read from that file and transferred through Internet, by the communication software, to the client where the operator, using the AR interface, establishes the position and orientation of each object.

Using this information a command is generated through the soft and transferred through Internet to the TIC, which command the telerobot, in order to execute the desired task.

IV. The Communication Software

The communication software technologies are based on PHP. A specific protocol over IP was designed for communication between the server-clients.

A. Server Interface

Application consists in two windows. The smaller is used to activate / deactivate the camera. Figure 3 presents a screen capture of the application.

The main window consists of three main regions:

- The upper zone, where control buttons (1, 2, 3 and ,4) can be found
- The window marked with 5 shows the image sent by camera, or, if inactive, the last image transmitted by the camera
- Zone 8 is a list of messages (client connection, data transmission, and so on)

- Zone 7 presents the 3D scene where the robot movement takes place.

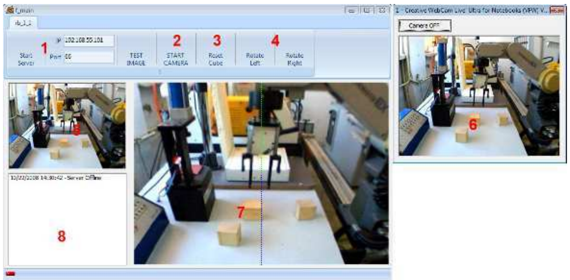


Fig.3. Server interface

Explanations of the main window zones:

1. The first zone (1) presents the server settings. In the IP textbox, the IP address of the computer on which the server part of the application runs is shown. The client side of the application must use this address. In the Port text box any number between 1 and 65535 can be written; this number has no significance for the client side. Start Server button is used to activate the application thus accepting income connections from clients.
2. The button Start Camera (2) activates window 6. This window contains a command button and an image. The command button is used to start / stop the camera. When camera is started, the image is also visible in the frame 5.
3. Command button (3) resets the cube in the scene (7) at its initial state (zero position).
4. The buttons in zone 4 allow the rotation of the cube to the left or to the right
5. The frame 5 contains the image delivered by the camera when the camera is activated or the last transmitted image if the camera is stopped.
6. In the frame 6, the image delivered by the camera is shown. If the camera is unavailable, this frame is empty.
7. Frame 7 is used to command the robot arm, positioning the green cube inside the image, using the mouse. Cube movement and control inside the frame is made using TGLCustomSceneObject OpenGL class objects.
8. Inside frame 8, information about server state, and information about image transfer and clients' state is provided.

The following code presents the command successions to process client commands, written in Delphi:

```
 s:=uppercase(AContext.Connection.IOHandler.ReadLn);
    sCommand := Copy(s, 1, 3);
    if ( sCommand = 'LST' ) then
      begin

AContext.Connection.IOHandler.WriteLn('LIST');
        AContext.Connection.Disconnect;
      end
    else
       if ( sCommand = 'CAP' ) then
         begin
           if ( captureIMG ) then
             begin
               if ( msStream.Size > 0 ) then
                 begin

AContext.Connection.IOHandler.WriteLn(
IntToStr(msStream.Size) );

AContext.Connection.IOHandler.WriteBufferOpen;

AContext.Connection.IOHandler.Write(msStream);

AContext.Connection.IOHandler.WriteBufferClose;
                  FreeAndNil(msStream);
                  lstRequests.Items.Add(
DateTimeToStr(Now) + ' - Image Transfer OK' );
                 end;
             end;
           AContext.Connection.Disconnect;
         end
     else
       if ( sCommand <> 'LST' ) AND ( sCommand <>
'CAP' ) then
         begin
           AContext.Connection.IOHandler.WriteLn('ERR
- Command');
           AContext.Connection.Disconnect;
         end;
```

Variable 's' represents the text provided by the client. It is verified for valid commands. In the above list, two commands (LST and CAP) are used.

LST command sends back the text "LIST" to the client, and then interrupts the connection.

CAP command sends first the file size then the image file and at the end of transaction interrupts the connection.

B. *Client Interface*

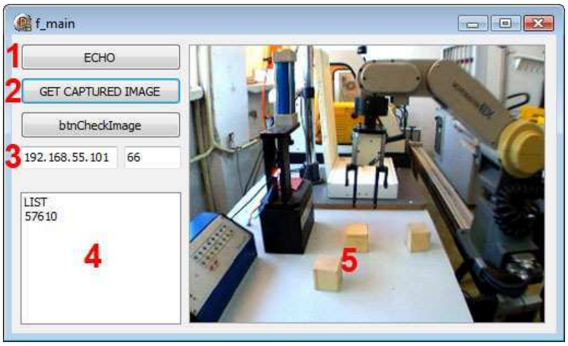


Fig.4. Client interface

Figure 4 presents a screen capture of the client application.

Command Button 1 (ECHO) is used to send LST command to the server.

Command Button 2 (GET CAPTURED IMAGE) sends CAP command.

Textbox (3) contains the server IP address and port.

Text area (4) shows messages sent by the server (When LST command is sent "LIST" text is received from the server)

Frame 5 contains the image sent by the server.

When command button 1 is activated, the following code is executed:

```
        Connect;
        IOHandler.WriteLn('LST');
        lstServerMessages.Items.Append(
IOHandler.ReadLn );
        Disconnect;
```

In line 2 LST command is sent to the server. The next function always processes data sent by the server. Line 3 shows "LIST" text contents sent by the server, in text area (4).

V. THE AR INTERFACE

The AR interface has been realized using Matlab software.

In the first step, the calibration of the system was made in order to improve accuracy and usability of the AR Interface.

The system calibration consists in two stages.

The first one was the camera calibration and this is made at the beginning when the system software is launched. To solve this problem we use the Devernay and Faugeras' technique [4] for lens distortion removal from structured scenes. The acquired image of the grid is presented in figure 5.

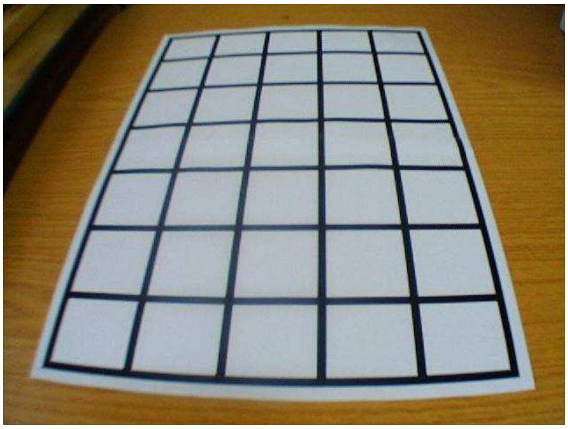


Fig. 5. The acquired image of the grid

The algorithm consists in the following steps:

- edge extraction on the acquired image as is presented in figure 6;
- the polygonal approximation with a large tolerance on these edges to extract possible lines from the sequence;
- finding the parameters of the distortion model that best transform these edges to segments;
- generating the "undistorted image" using the parameters computed using this algorithm (k - radial distortion term. c_x, c_y - x and y coordinates of lens centre expressed as fraction of the image size relative to the top left corner; s - apparent aspect ratio.)

The next step is the calibration of the interfacing image. The purpose of this module is to map the two-dimensional coordinates as shown on the captured image to three-dimensional coordinates in real space using four reference signs. The algorithm that simulates the third coordinate dimension (depth) is based on a single vanishing point model.

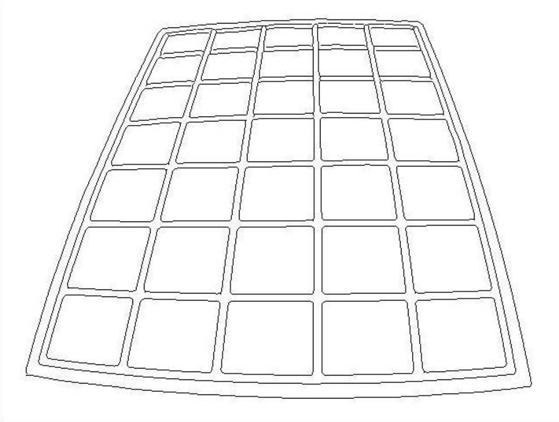


Fig. 6. The edge extraction on the acquired image

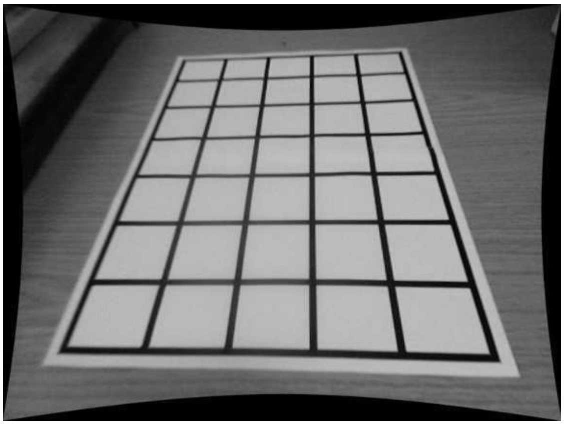


Fig. 7. The "undistorted image"

The method consists on the superposing of the "virtual reality" image and the real scene in which robot acts. The real image can be re-created in virtual space in two steps:

1. Manual superposing of the viewing direction of the real camera over the viewing direction of the virtual camera (i.e. the biunique correspondence of the angles under the scene is viewed both in real and virtual reality)
2. This step contains two cases:
   a. Either manual superposing of the distances between the real camera and the real scene with the distance between virtual camera and virtual scene (the movement of the real camera), or
   b. The movement of the virtual camera to attain the same distance between the camera and the scene as in the real world.

At this moment the scene calibration can be made using the four corners of the scene boundary as the four reference signs mentioned above. Thus, a point-to-point correspondence between real scene and the virtual one is made, and the system is calibrated. Calibration is made the very first time the system is turned on, calculating in Matlab® the correction coefficients used in the main program.

After the calibration, for each type of the real object in the scene a wireframe model is generated using geometrical primitives. Using 3D transformations (translation, rotation and scaling) wireframe models can now be moved at the desired location.

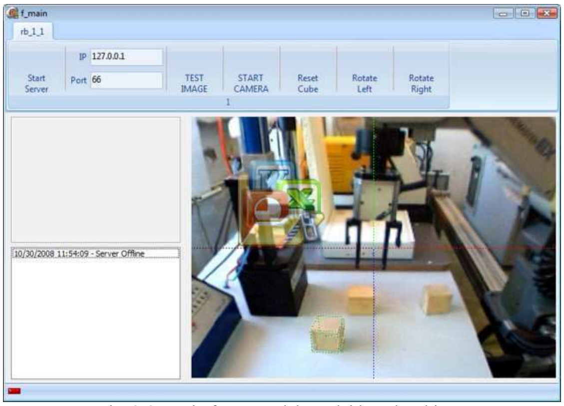


Fig. 8 One wireframe model overlaid on the objects

The dimensions of the object model in the robot's image plan are computed through 3D to 2D transformations, considering the vanish point, thus resulting the object's model in the image plane which is overlaid on the object's image (in the image plane – figure 8).

The mass centre position and the orientation of the object are computed through a software procedure and are used to command the robot.

VI. THE ROBOT CONTROL

Having that information the human operator transfers via Internet a command to the remote computer; this transfers it to the robot controller through parallel port. The telerobot will execute the task.

VII. CONCLUSION

The Mitsubishi Telerobot project demonstrates how much an improved AR interface can increase the performance of a telerobotic system without having to change any of the telerobot technical features.

The next step in development of our telerobot system is to include in the AR interface not only the visual sense, but also the haptic, using haptic gloves and HMD to command and control the process.

The project was successful in the development of the AR interface for the Mitsubishi Telerobot. The objective of the project was therefore met.

VIII. REFERENCES

[1] Albeanu G. *Quality indicators and metrics for capability and maturity in E-learning*, The 3rd Int. Sci. Conference, "E-learning and Software for Education", Bucharest, April 12-13, 2007, pp. 165-175.

[2] R. T. Azuma, *A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments* 6, 4 (August 1997), 355-385.

[3] Chen S.H., Chen R. et al., *Development of Remote Laboratory Experimentation through Internet*, IEEE Symposium of Robotics and Control, Hong-Kong, July 1999, pp. 756-760.

[4] F. Devernay and O. Faugeras. *Automatic calibration and removal of distortion from scenes of structured environments*. In SPIE, volume 2567, San Diego, CA, July 1995

[5] Fellner D.W., Hopp A., *VR-LAB - A distributed multi-user environment for educational purposes and presentations*, Technical Report TUBS-CG-1999-03, TU-Braunschweig, 1999.

[6] M. Fezani, C. Batouche, A. Benhocine, *Study and Realization of the Basic Methods of the Calibration in Stereopsis For Augmented Reality*, American Journal of Applied Sciences 4 (3): 297-303, 2007

[7] H. Friz, (1999) *Design of an Augmented Reality User Interface for an Internet based Telerobot using Multiple Monoscopic Views*. Diploma Thesis, Institute for Process and Production Control Techniques, Technical University of Clausthal, Clausthal-Zellerfeld, Germany Available at: http://telerobot.mech.uwa.edu.au

[8] K. H. Y. Fung, B. A. MacDonald and T. H. J. Collett, *Measuring and improving the accuracy of ARDev using a square grid*, www.araa.asn.au/acra/acra2006/papers/paper_5_45.pdf

[9] Gibson, S., Cook, J., Howard, T., Hubbold, R., Oram, D.: *Accurate camera calibration for off-line, video-based augmented reality*. In: IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2002), Darmstadt, Germany (2002)

[10] Han K.-H., Kim S. et al, *Implementation of Internet-based personal robot with internet control architecture*, IEEE Int. Conf. on Robotics & Automation, Seoul, Korea, May 21-26, 2001, pp. 217-222.

[11] F. Harald *Design of an Augmented Reality User Interface for an Internet based Telerobot using Multiple Monoscopic Views*, Diplomarbeit, Institute for Process and Production Control Techniques, Technical University of Clausthal Clausthal-Zellerfeld, Germany

[12] Hazry D., Sugisaka M., Yuji T., *Human-Robot interface over the web based intelligent system*, American Journal of Applied Sciences, vol. 3, no. 1, 2006, pp. 1634-1639.

[13] K. Jong-Sung, H. Ki-Sang, A recursive camera resectioning technique for off-line video-based augmented reality, Pattern Recognition Letters 28 (2007), pp. 842–853

[14] G. Klancar, M. Kristan, R. Karba, *Wide-angle camera distortions and non-uniform illumination in mobile robot tracking*, Robotics and Autonomous Systems 46 (2004), pp.125–133.

[15] T. H. Kolbe, *Augmented Videos and Panoramas for Pedestrian Navigation*, Proceedings of the 2nd Symposium on Location Based Services & TeleCartography, 2004, Vienna

[16] Kosuge K., Kikuchi J., Takeo. K., *VISIT: A Teleoperation system via computer Network*, Proc IROS'98: Workshop on Web Robots, Victoria, Canada, l2-17 0ctober, 1998, pp. 61-66.

[17] Langmann R., *The web-controlled mobile communication Robot-Ego Secundus*, International Journal of Online Engineering, Vol. 3, No. 1, 2007, http://www.i-joe.org/ojs/viewarticle.php?id=84.

[18] Liao C., Liu Q., et al., *A tele-robot assistant for remote environment management*, IEEE Int. Conf. On ICME, Vol. 3, 2004, pp. 1827-1830.

[19] Lobov A., Lastra J.L.M., Tuokko R., *A collaborative framework for learning robot mechanics: RIO-robotics illustrative software*, The 33rd ASEE/IEEE Frontiers in Education Conference, Boulder, november 5-8, 2003, pp. F4E 12-16.

[20] Matijasevic M., Valavanis K.P. et al., *Application of a multi-user distributed environment framework to mobile robot teleoperation over the Internet*, Machine Intelligence & Robotic Control, Vol. 1, No. 1, 1999, pp. 11-26.

[21] Nini, B., Batouche, M. *Utilisation D'une Sequence Pour L'augmentation En Realite Augmentee*, www.irit.fr/recherches/sirv/congres/jig05/nini.pdf

[22] Palmer, R. *Augmented reality and Telerobots*, Honours thesis, University of Western Australia.

[23] Piguet Y., Gillet D. et al., *Environment for online experimentation and analysis*, IEEE Int. Conf. on Robotics and Automation, Workshop on Educational Applications of Online Robots, Washington DC, May 11-15, 2002.

[24] Scultz D., Burgard W. et al., *Web interfaces for mobile robots in public places*, IEEE Magazine on Robotics and Automation, Vol. 7, No. 1, 2000, pp. 48-56.

[25] Scultz D., Burgard W., *Robust visualization for web-based control mobile robots,* In Goldberg K. and Siegwart R., (eds.), Beyond Webcams, MIT Press, 2001.
[26] T. B. Sheridan, *Telerobotics, automation and human supervisory control*.Cambridge, MA: MIT Press.
[27] H. Schaefer, *Kalibrierungen für Augmented Reality*, Diplomarbeit, Fachgebiet Photogrammetrie und Fernerkundung Technische Universität Berlin, November 2003.
[28] Thamma R., Huang L.H. et al., *Controling robot through Internet using Java*, Journal of Industrial Technology, Vol. 20, No. 3, 2004, pp 2-6.
[29] R. Y. Tsai. *A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf TV cameras and lenses*. Robotics and Automation, IEEE Journal of, 3(4), Aug 1987.
[30] Tuttas J., Wagner, B, *Distributed Laboratories*, Proc. Int.. Conf. on Engineering Education, August 6-10, 2001, Oslo, Norway
[31] Weaver A. C., Luo J., Zhang X., *Monitoring and control using the Internet and Java*, IEEE Proc. Int. Conf. Industrial Electronics, San. Jose, CA, 1999, pp. 1152-1157.
[32] Yang X., Chen Q. et al., *Internet-based tele-operation of a robot manipulator for education*, Proc. IEEE Workshop on Haptic Audio Visual Environments and their Applications, Ottawa, Canada, October 2004, pp. 7-11.

Simultaneous Use of Scaled and Virtual Models in Early Architectural Design Education

Neander F. Silva, Ecilamar M. Lima
Universidade de Brasília
Laboratório de Estudos Computacionais em Projeto – LECOMP
ICC Norte, Ala A, Subsolo, CEP 70910-900, Brasília, DF, BRASIL
http://lecomp.fau.unb.br/, neander@unb.br, ecilamar@unb.br

Abstract – We present in this paper a teaching experience in a design studio course in an architecture school which had the objective of dealing with pedagogical difficulties in integrating computers and its most advanced tridimensional modeling capabilities into the design studio. 3D computer modeling was introduced simultaneously with scaled modeling into at an early design studio of our curriculum. Both types of representation (virtual and scaled) were mandatory. Our objective was to challenge the widespread idea among many architecture teachers and students that computers should not be taught before than or simultaneously to the teaching of architectural hand drawing since this would lead to less creativity, poorer products and increased repetition of solutions. In our experience we found out that the students did manage to design by representing 3D artifacts since the conception, there was no evidence of less creativity or repetition and there was no evidence of any inability for hand sketching.

I. The Research Problem

The introduction of computers has brought many changes to architectural offices in the last 20 years. Initially it was just mimicking the traditional desktop through the automation of drafting and construction drawings. However, the work was still done two-dimensionally producing floor plans, cross sections and elevations, without allowing integrated visualization of the parts of the design project. Even today, as stated by Stacey, Beesley and Hui in [1]: "…two-dimensional drafting still dominates the construction industry… The revolutionary potential of three-dimensional modeling is used fitfully… only by a few."

We believe this underuse of computer's potential is related to the educational approach adopted by most architecture schools because the professionals formed by them continue to reproduce the design method they learned. Many of the architecture schools do not emphasized three-dimensional (3D) representations in their design studios. Some have argued in the past that 3D representations are too complex and labor intensive and 2D representations allows for the reduction of such complexity in the design process [2]. However, other authors, such as Zevi [3], since long time ago, have argued that architecture cannot be apprehended through 2D representations. Floor plans, section and elevations give a partial view of the building and their sum does not provide a complete view of the whole project. It is necessary to introduce time as a fourth dimension to fully understand a building only as we enter it and move around within it.

Zevi argued that the traditional 2D drawings are not only fragmented and misleading representations of architecture, but he has also pointed that they were never intended to be used as design media. They were invented and developed just as construction instructions to those who are suppose to build what was designed.

Many have tried to overcome the representational problem using three-dimensional scaled models. This strategy can be observed in professionals as diverse in time and architectural style as Antonio Gaudi [4] and Frank Gehry [5]. A scaled model allows the conceptual design to take place in a holistic way, controlling many variables simultaneously. If adequately built with materials in the right size for the chosen scale, it allows the testing of constructive and structural problems as it can be seen in the work of Gaudi [6].

However, the scaled model is not a complete solution in terms of representing architecture. Although it can represent all its dimensions, it suffers from the same limitations of permeability of a sculpture: it can only be observed from the outside. Besides, it also suffers from the very limitations of being scaled. As an example, we usually see the scaled model from above, as in an aerial view. This situation is very unlikely to happen in the real world. We usually apprehend the building from the ground in an infinite sequence of perspective views as we walk in and along [3].

In the context of the technological development of 1948, Zevi could not conceive a way in which architecture could be properly represented in its four dimensions [3]. However, since mid 1980's, interactive three-dimensional computer modeling provides the answer for architectural representation, as it is well known from the literature [7].

A virtual three-dimensional model is not just an equivalent to a scaled model as it has been believed by some in the past [8]. Besides representing the four dimensions of architecture,

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_99, © Springer Science+Business Media B.V. 2010

it allows for many possibilities which cannot be provided by a scaled model or any other traditional media. A virtual 3D model offers many exclusive resources such as interactive walk-through, object animation, solar animation, global illumination calculation, designing complex non Euclidian geometries [9], digital fabrication [10] and mass customization [11].

We present in this paper a teaching experience in a design studio course in an architecture school which had the objective of dealing with pedagogical difficulties in integrating computers and its most advanced tridimensional modeling capabilities into the design studio.

Firstly, the school curriculum had no experience in integrating computers into the design process. Computers were not used in the design studio except for documentation through the production of construction drawings.

Secondly, the school emphasized the teaching of computer based on two-dimensional documentation and did not give room for tridimensional computer modeling (3D modeling) in the mandatory curriculum. There was only one computer related course for teaching 2D computer-aided drafting (2D CAD) in the second half of the first year. In this course the students were taught to represent existing buildings, instead of designing new ones.

Thirdly, the most notably consequences were these: our curriculum was considerably outdated in comparison with other schools which embedded the teaching of 3D modeling into early design studios for at least 15 years [7]. Also, the method of designing from documenting their projects in multiple 2D representations, rather than representing the artifacts being designed, had become much engrained into the students' minds. An elective course, with the purpose of teaching 3D modeling in a design context, located later in the curriculum, has proven very ineffective in changing such rooted practices.

II. Hypothesis

Our hypothesis was: 3D modeling could be introduced simultaneously with scaled modeling into the design studio before 2D CAD. In this way we would be able to overcome the existing limitations of design carried out with misleading two-dimensional representations that uses computers only as presentation. Instead we would teach computers as design tools since the project's conception.

Our objective was to challenge the widespread idea among many architecture teachers and students that design computer media should not be taught before than or simultaneously to the teaching of architectural hand drawing since this would lead to less creativity, poorer products and increased repetition of solutions.

III. Research Method

We have adopted in our research project a problem-based learning approach in which we embedded the teaching of three-dimensional computer modeling directly into the design studio. The objective is to encourage students to learn how to use 3D modeling as design tool in the decision making process and not just as a drawing tool [12][13][14].

The course we are responsible for is called *Architectural Design Studio 2 – Language and Expression* and is located in the second half of the first year. The emphasis of this course is on form exploration and creation and on the use of architectural languages and vocabularies in this process.

Among other objectives of this course we could cite: 1) the introduction of the students to the formal aspects of the architectural project, observing their relations with the functional, environmental and constructive building problems; 2) the introduction to the study of the form, space and principles that guide the organization of the building; 3) the contribution for the acquisition of a repertoire through the analysis of design precedents; 4) the introduction to the design practice through the integrated representation, visualization, analysis and synthesis of the artifacts being designed.

Two small design projects were developed and a scaled model and a virtual 3D model were required for each of them. We have adopted as themes for this course, in the last two terms, the design of a gateway or arch and the design of an exhibition pavilion in the context of an entertainment area. The students had three weeks to conceive and develop the first project and six weeks for the second one.

The verification of our hypothesis was based on the method of designing by thinking the architectural object as a whole, by using primarily 3D representations since the conceptual design stage. This meant using scaled models and 3D virtual models simultaneously. Freedom was given to students to choose on what media to begin the project: the virtual or the scaled model. However, the projects had to be conceived and developed in these two media. There were 28 students registered in the course.

IV. Results

The results were assessed against the following criteria:

1) Did the students manage to design primarily by representing 3D artifacts in virtual and scaled models? The answer to this question is yes because the all the students produced a model in both medias. Besides, many of them produced also intermediate prototypes for study purposes.

2) Did they resort very often to 2D representations? The answer is no. Only few of them resorted to 2D representations, but among those several made an effort to use perspectives rather than orthographic projections. The majority, about 75%, experimented with the 3D resources.

3) Was there any evidence of less creativity, poorer products and increased repetition of solutions? The answer is no. There was a large degree of creativity and diverse forms. Only about 14% of the students work was based on standard rectangular shapes. 75% of the students resorted to curved shapes in their projects. There is no resemblance with known precedents or with other works in the same sample in 60% of design projects. In addition 86% of the projects were structurally sound. The students strived to be venturesome basing their projects on non-standardized components that would require digital fabrication in 64% of the cases. Figure 1 contains a sample of the students' projects which validates our hypothesis.

4) Was there any evidence that the use of computers led to greater inability for hand sketching? The answer is no. It was just about the learning of how to use the virtual tool and media for the design project.

Above all the students were able to overcome the limitations of the 2D representations and all of them learned how to use the three-dimensional modeling, gaining an integrated perception of the project as a whole, very close to the experience of buildings in the real world.

V. CONCLUSIONS

We believe the originality of our contribution resides in the emphasis on simultaneous use of scaled and virtual models as design conceptualization media in early architectural design education.

Some similar experiences come close to ours, but they all differ in a number of ways. It might be argued that designing with scaled models is not new. Gaudi did it. However, it was simultaneously with other traditional design representations and, obviously, not with virtual models.

It might also be said that Frank Gehry designs using scaled models and virtual models simultaneously, but that is not quite the true. In his office the production of the virtual model only starts when the conceptual design is already finished in a scaled model [5].

The experience of the School of Architecture at Penn State University, to the best of our knowledge, places emphasis on 3D computer modeling and it does not emphasize simultaneous use of scaled models.

We believe our research showed promising results: we found out that the students did manage to design by representing 3D artifacts since the conception, there was no evidence of less creativity or repetition and there was no evidence of any inability for hand sketching. In fact our students projects showed a large degree of creativity through new and complex forms, no fear of experimentation and a venturesome attitude.

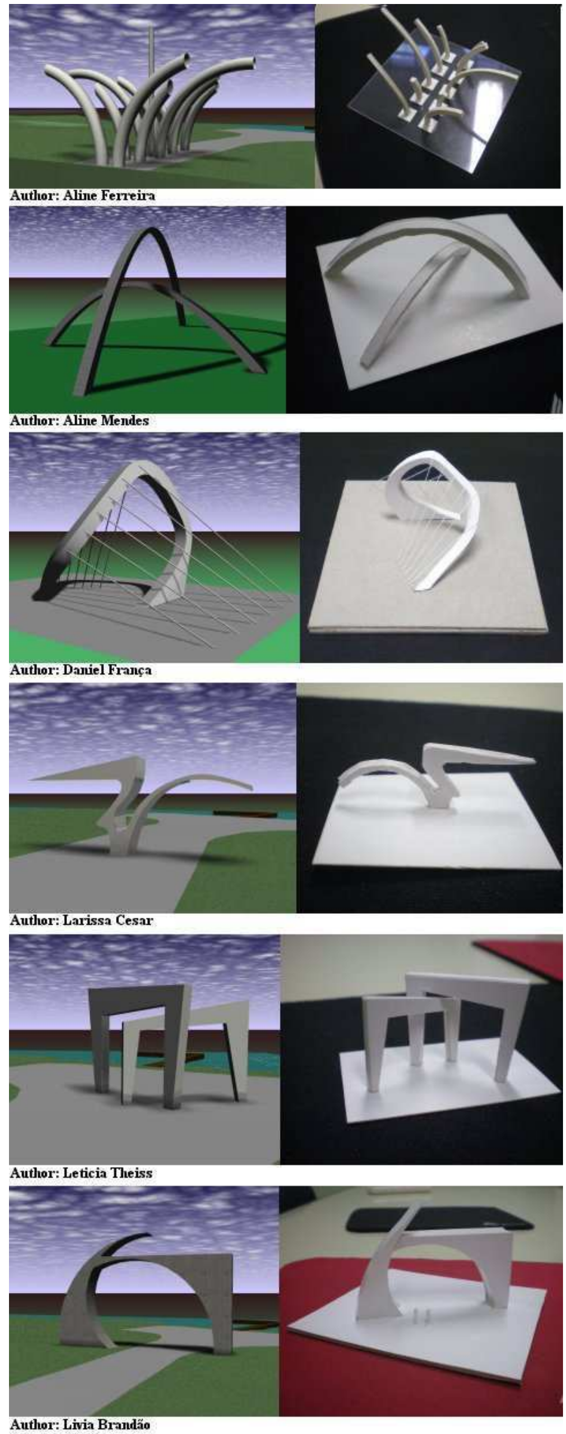


Fig. 1. Virtual (left) and scaled models (right).

REFERENCES

[1] Michael Stacey, Philip Beesley & Vincent Hui, “Digital Fabricators”, University of Waterloo School of Architecture Press, p. 6, 2004.

[2] Alexander Koutamanis, “The future of visual representations in architecture”, in Automation in Construction, vol. 2, number 1, 1993.

[3] Bruno Zevi, “Architecture as space – How to look at architecture”, 1st American edition, Horizon Press, New York, pp. 22, 23, 26, 1957.

[4] Mark Burry, “Between intuition and process: parametric design and rapid prototype”, in Branko Kolarevic (editor), “Architecture in the Digital Age – Design and Manufacturing”, Taylor & Francis, New York, pp. 151, 156, 2003.

[5] Sydney Pollack, “Sketches of Frank Gehry”, video, Thirteen/WNET New York’s American Masters, 2006.

[6] Branko Kolarevic, “Architecture in the Digital Age – Design and Manufacturing”, Taylor & Francis, New York, p. 5, 2003.

[7] Loukas Kalisperis, “CAD in Education: Penn State University”, in ACADIA Quarterly, vol. 15, number 3, 1996.

[8] Brian Lawson, “Design in Mind”, Architectural Press, Oxford, p. 145, 1994.

[9] Branko Kolarevic, “Digital Morphogenesis”, in Branko Kolarevic, “Architecture in the Digital Age – Design and Manufacturing”, Taylor & Francis, New York, p. 13-28.

[10] Branko Kolarevic, “Digital Production”, in Branko Kolarevic, “Architecture in the Digital Age – Design and Manufacturing”, Taylor & Francis, New York, p. 31-51.

[11] Branko Kolarevic, “Digital Production”, in Branko Kolarevic, “Architecture in the Digital Age – Design and Manufacturing”, Taylor & Francis, New York, p. 52-53.

[12] Barry Maitland, “Problem-based learning for architecture and construction management”, in D. Boud and G. Feletti (editors) “The Challenge of Problem-based Learning”, Kogan Page Ltd, London, pp. 211-217, 1997.

[13] Alan Bridges, “A Critical Review of Problem Based Learning in Architectural Education”, in Communicating Space(s), in the 24th eCAADe, Conference Proceedings, Volos, Greece, pp. 182-189, 2006.

[14] Neander Silva & Ecilamar Lima, “Design Computing Education: A comparative study among learning environments”, in Magued Iskander (editor) “Innovative Techniques in Instruction Technology, E-Learning, E-Assessment and Education”, Springer, New York, pp. 387-391, 2008.

Modeling A Deburring Process, Using DELMIA V5®

Mohammad Mohammad, Vipul Babriya, Tarek Sobh
School of Engineering
University of Bridgeport,
Bridgeport, CT 06604 USA

ABSTRACT: **This paper investigates a Deburring process which is integrated with a robotics application. Conventional tumble deburring processes have been used for years in the surface finishing industries. The manual process is tedious, inconsistent and inaccurate. Automation in the finishing process has been proved very beneficial. Integration of robots in the process provides more flexible and convenient process. We investigate the off-line simulation of finishing edges using the Product Life Cycle tool DELMIA®. The process extracts the feature to generate a part finishing process plan. Collisions with the environment are detected and avoided through the simulation of cutting the path before the path is downloaded to the actual hardware. Either the tool or the part is being fixed on the robot's end effector and simulated with respect to the defined path. This paper presents the complete off-line simulation of the Part in Hand as well as Tool in Hand approaches for the deburring process**.

KEYWORDS: Dexterity, Offline Simulation, Flexi Burr, Tool Orientation, Manufacturing, Robotics.

I. INTRODUCTION

As the number of robotic applications and installations continue to grow, the need for effective and efficient robot programming techniques grows proportionally. In the past, almost all robotic involvement was through teach-by-show techniques/manual deburring operations. This implied that the robot's paths are generated either by mounting a dowel pin that matches the tool's diameter or by mounting a pointed teaching tool on the robot. The pin or the teaching tool is moved manually to a point where they touch a finished part edges, and the software records that point. After repeating this process along the part edges, the robot controller uses the recorded points to define the path. The more complex the part, the more programming time is required to achieve an acceptable path [1].

A. *Off-line simulation*

Off-line simulation refers to the technique of running models in a purely stand-alone format, not connected to any hardware and normally not run in Real-Time. Off-line simulation and lead-though programming promise acceleration and streamlining of the robot programming process [2]. They enable the user to program robot motions in a simulated "virtual computer environment".

B. Tumble deburring operation

Tumble Deburring removes sharp edges or burrs, and smoothes surfaces of metal components [resulting from some manufacturing or casting process]. It can be done as a wet or dry process. A six- or eight-sided machine with either horizontal or vertical axis is used with abrasive "stones".

The deburring tool could be small, and a part with many small/accurate details is hard to be reached via a CNC machine. A good solution would be to use an industrial robot (with 6 axis) to perform the deburring process. Pratt & Whitney has estimated that 12% of their total machining hours are devoted to manual deburring and chamfering of parts after they have been machined [3].

II. BASIC CONFIGURATIONS OF DEBURRING PROCESS PERFORMED BY INDUSTRIAL ROBOTS

A. *Tool in Hand*

In these applications a compliant tool is mounted on the robot end effectors and manipulated over the part to be finished. Tool in Hand configurations are used where the part to be finished is too large or unwieldy for a robot to carry. Belt media is rather rare in tool in Hand applications because it is difficult to build compact tools using belts. Figure 1 illustrates a tool in hand configuration.

B. *Part in Hand*

The work piece is fixed on the robot end effector while the tool is fixed as an auxiliary device and the robot manipulate the work piece to be deburred as result of contact between the work piece and the tool. Part in hand applications are most often used when the part to be finished is relatively small in size (comparing with deburring/grinding/polishing tool's size) .Gripper tooling allows the robot to pick up the part and manipulate it against the abrasive finishing media [4]. Figure 2 illustrates a part in hand configuration.

2.2.1 Advantages of the Part in Hand approach

- Often, robot load/unload operations can be combined with surface finishing operations on a single work station; i.e., a robot can remove a part from a serial line conveyor, finish the part, and then place the part in final packaging. Doubling up these operations can provide a much greater return on investment.
- The surface finishing apparatus, whether it is a belt, wheel or disk device, can be quite large using longer belts, large diameter wheels, and higher horsepower, meaning that parts can be processed faster with longer intervals between media changes.

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8_100, © Springer Science+Business Media B.V. 2010

2.2.2 Disadvantages

Sometimes it is impossible to finish the entire surface of the part. This can be due to interference with the robot gripper and insufficient robot dexterity to reach around the part in different positions.

Fig.1 Tool in hand configuration

Fig. 2 Part in hand configuration

III. Examples of Industrial Robots Used in Deburring Processes

A Motoman

Figure 3 shows the Motomon DX1350N robot. It is frequently used in deburring industrial application. Following is the configuration of the robot.

Motoman DX1350N
Max Reach: 1355 mm
Max Payload: 35kg
Weight: 27kg
Repeatability: 0.6mm

Fig. 3 Motoman DX1350N

B. Nachi

Figure 4 illustrates the Nachi SC35F robot. The Nachi SC35F industrial robotic system is designed for heavy loads and long strokes, with high positioning repeatability and speed [5]. The SC35F is a lightweight design allowing for easy installation and transport. Following is the configuration of this robot model.

Nachi SC35F
Max Reach: 2002mm
Max Payload: 35kg
Weight: 400kg
Repeatability: 0.1mm

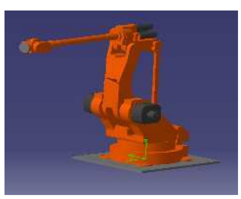


Fig. 4 Nachi SC35F

C. ABB

The ABB IRB 140 is most commonly used in surface finishing industries. This robot is a compact and powerful 6-axes machine with a unique combination of fast acceleration, large working area and high payload [5]. Figure 5 shows the IRB 140. Following is the configuration of the robot.

IRB 140
Axis: 6
Reach: 810 mm
Payload: 5kg
Repeatability: 0.03 mm

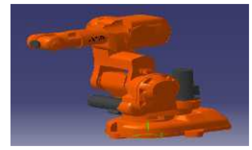


Fig. 5 ABB IRB 140

D. Kuka

The exclusive use of stainless steel on all surfaces, together with high IP rating, makes the KR 15 SL suitable for fields of applications with stringent requirements as to hygiene, sterility and absence of particles – such as in the food handling industry or medicine as well as the surface finishing industry [5]. Figure 6 illustrates the Kuka KR 15 SL model.

Axis: 6
Reach: 1503 mm
Payload: 15 kg
Weight: 315 kg
Repeatability: .01 mm

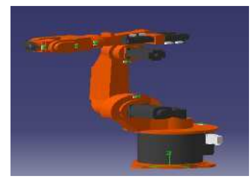


Fig. 6 Kuka KR 15 SL

IV. Tool Used in Deburring Processes

The success of most automated deburring operations is dependent on the flexibility and consistency of the deburring tool. Problematic burrs, residual material on parting lines, and flashing on die-cast parts must be removed. Systems designers typically program the robot to move its deburring tool along a path defined by discrete points. That path, however, may not coincide exactly with the shape or contour of the surface to be deburred because of variations in the part itself or differences between the part edge and the exact path the robot has interpolated [6]. Otherwise, the system designer has to create thousands of very physically close tag points automatically to reduce the error. Figure 7 illustrates the interpolated path defined by off-line simulation and actual path/edge that the robot is supposed to follow.

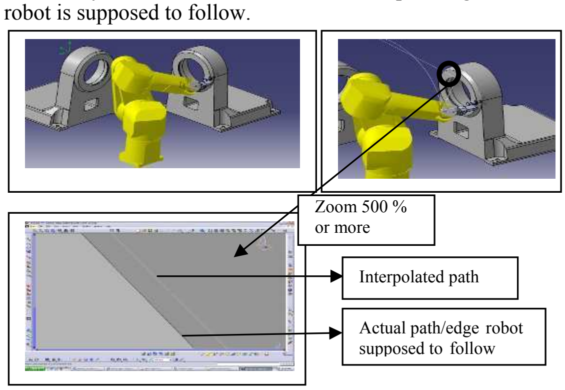


Fig. 7 Illustration of the interpolated and actual path of robot supposed to follow

A. Deburring tools with feedback force sensors [Flexdeburr]

Flexdeburr is a high-speed, air turbine driven tool for deburring aluminum, plastic, and steel [7]. While spinning at high speeds, the lightweight, rotary tool has radial compliance supported by air pressure applied to the shaft allowing the tool to perform consistently on irregular part patterns. Figure 8 illustrates, what occurs when a non-compliant tool is being linearly interpolated between taught points on a contoured surface. The surface contours in figure 8 are greatly exaggerated for clarity; however this effect exists in parts with even a slight contour. As shown, even though path points are taught perfectly on the surface, as the robot interpolates between the points, the grinding media position relative to the part surface varies. On actual parts with slighter contours, the varying contact can be detected by listening for variations in the grinding/deburring motor speed [8].

The bottom illustration in Figure 8 shows the same surface but using an auxiliary compliant tool. With this setup, it is possible to maintain consistent media contact with the surface with only a minimum number of taught points. In addition, if the path is linear or circular, it can be interpolated with a few number of discrete tag points. The problem always occurs with linear interpolation between tags on an irregular edge of the work piece. Thus, using compliant force tool design, one does not have to generate a large number of points connected by interpolating linear path, but the tool will correct the error in real time. Steps of the process on Delmia V5 will be clarified later in this paper.

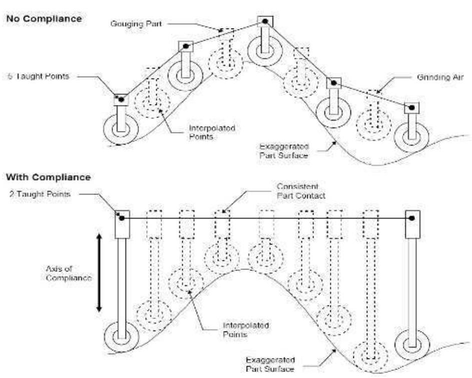


Fig. 8 Illustration of the compliance on surface contact

Figure 9 illustrates a deburring tool with feedback sensor. The deburring tool has a rigid outer housing and internal motor/spindle assembly that provides the compliance. The pneumatic motor/spindle assembly is mounted on a pivot bearing mounted to the tool housing. This allows the pneumatic motor to move with the pivot bearing independently of the housing. The radial "compliance field" is created by a ring of small pneumatic positions located near the front of the tool housing [9]. The compliance force can be exerted on the spindle/tool in any direction (360 degrees) radially from the tool.

Fig. 9 Flexdeburr

B. *Force/Torque sensor*

In the other types of deburring processes, when the robot holds the work piece and the tool/grinding belt is fixed as shown in figure 10. There should also be a fixed force/torque sensor between the robot and the work piece to maintain constant contact force in surface finishing application.

Fig. 10 Illustration of force/torque sensor in grinding process

C. Path Profile

Robot motion for both part in hand and Tool in hand applications involves smooth sweeping movements. There are three important aspects of the motion which have to be carefully considered in order to have a good finish:

1. Surface speed
2. Part/Tool Orientation
3. Approach Vector

4.3.1 Surface Speed

There are four process variables which most greatly affect the Material Removal Rate (MRR) in finishing operations. These variables are:

1. The aggressiveness of the media.
2. The force with which the media is applied.
3. Rate at which the media is fed (RPM).
4. Speed at which the media is moved over the part surface.

It should be noted, that all these variables can influence and offset each other when trying to achieve a particular surface finish while maintaining a desired process throughout. For example, a more aggressive media can be used with lighter applied force at lower RPM so that a higher surface speed can be used to increase production rates. However, this must be balanced with the fact that aggressive media does not produce extremely fine surface finishes. The one rule that should be remembered regarding surface speed is that the more aggressively a part being worked on a given pass (i.e., aggressive media at high force levels with high rpm) the faster the surface speeds. Thus, it can cause overheating the part surface and media.

Unfortunately, there is no theoretical approach to determine the optimal combination of the process variables. Trial and error combined with prior experience seem to be the best method to determine these variables.

4.3.2 Part/Tool Orientation:

The relative orientation of the part and tool is very important in achieving a consistent, uniform surface finish. The overall surface appearance is directly affected by the consistency of location and shape of the area where the finishing media contacts the part. This area is called the contact patch. The contact path is "where the rubber hits the road". Therefore, the single most important path programming consideration is to maintain constant contact patch, no matter what configuration is used, Part in Hand or Tool in Hand, and no matter what programming method is used. In most finishing operations, it is usually desirable to have the finishing media applied to the part surface at an angle varying from normal to the surface to three or five degrees off the normal direction depending on the type of media.

4.3.3 Approach Vector

In the approach vector the part comes first with the finishing media. This transition point often is where many surface finishing inconsistencies are found in the final part. The first rule that applies to the approach vector is to approach gradually. If one looks at how a person finishes parts, one would find that person's movements are characterized by sweeping motions with gradual, controlled setting down and lifting of the media. These gradual and controlled motions serve to "feather" the finish at the edges.

Active tools, with closed-loop control, have the unique ability to eliminate these undesirable effects by giving robots a more human touch [10]. These tools are able to automatically compensate for stiction and inertia effects in the same way as human operators anticipate these effects.

4.3.4 Tool Example:

Figure 11 illustrates an active tool, with closed- loop control. The filing tool SWINGFILE 2000 is well suited to remove burrs particularly in narrow slots and grooves. It can be employed for metallic as well as nonmetallic parts. It is also well suited for the deflashing of aluminium die casting. It can be employed with any orientation. The file tip is compliable in two directions. The pressure against the part edge can be set from the robot program.

The filing tool can be mounted stationary either onto a tool stand or onto the robot arm. If these tools have to be changed automatically, a tool changer is available. Commercial filing insertion can be used. The air motor needs oiled compressed air for optimal life.

Fig. 11 Active control tool (SWINGFILE 2000)

V. Steps / terminology for deburring [tool in hand/part in hand] via delmia v5 r17

Figure 12 illustrates the block diagram of a complete procedure for deburring [Tool in Hand/Part in Hand] via DELMIA V5 R17 [11].

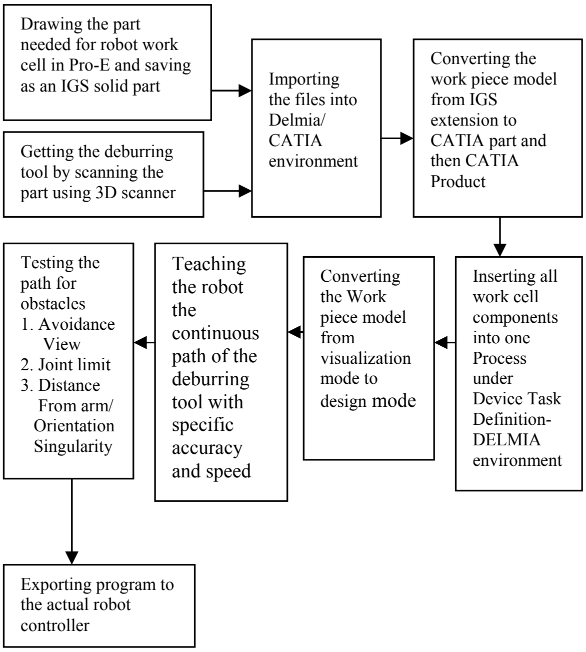


Fig. 12 Hierarchical block diagram of deburring process via DELMIA V5R17

VI. Steps for Tool in Hand Approach Using DELMIA V5R17

6.1 Inserting the all components of the process in Device task definition.
6.2 Attaching the tool to the robot.
6.3 Adjusting the positions of every component.
6.4 Adding new task to the robot program.
6.5 Automatic generating tag points.
6.6 Simulating for testing the scenario/process.
6.7 Adjusting & interpolating the tag points.

Figure 13 illustrates a snapshot of the robot, tool and part in the DELMIA environment.

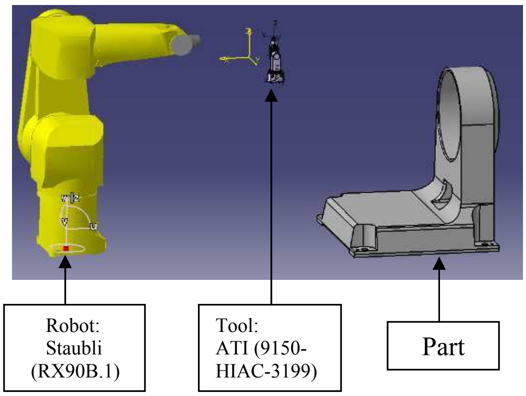


Fig. 13 Illustration of robot, tool and part in a DELMIA environment

A. *Procedure in Details*

6.1 Inserting all components of the process in device task definition.

Figure 14 illustrates how to start with device building in DELMIA.

1. File – open "basesolid.igs".
2. Save as basesolid.cat part.
3. Start Resource – Device Building.

Figure 15 illustrates procedure to generate a CATIA product.

Fig. 14 Snapshot of device building

Fig. 15 Screenshot window of CATIA product

4. Right click on the P.P.R Tree on Application →Existing component → Choose the Part name "basesolid.cat part"
5. Save it as basesolid.cat Product
6. Start-Resource – Device Task Definition

Figure 16 illustrates the start up window of device task definition.

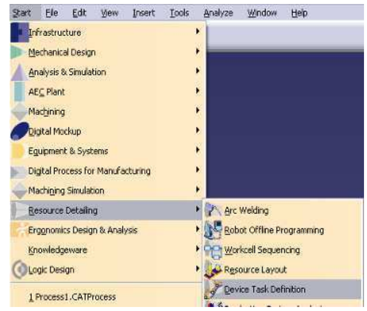


Fig. 16 Snapshot of start up device task definition in DELMIA.

7. Insert Product 
8. Choose part name and tool name
9. Insert Resource 

Path: C:\Program Files\Dessault/Systems\B17\intel_a\startup\Robotlib\V5DEVI-CES

6.2 Attaching the tool to the robot.

1. Select Attach the tool icon  from robot management toolbar.

A Robot Dressup window appears on the screen as depicted in figure 17.

- Choose the robot first.
- Choose the tool second.
- Change the snap reference to Base 1.
- Change Tool Centre Point (TCP) to Tool 1.

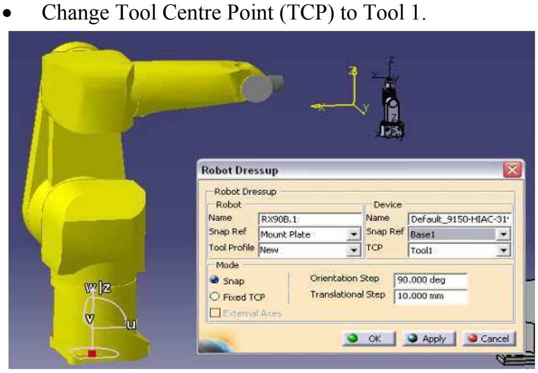
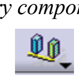


Fig. 17 Snapshots of robot and tool with dress up window in DELMIA

6.3 Adjusting the positions of every component

1. Align Side (optional procedure) 

- Choose the down side of the part (as master).
- Choose the down side of the robot (slave).

The robot base will be in same plane with as the base as shown in figure 18.

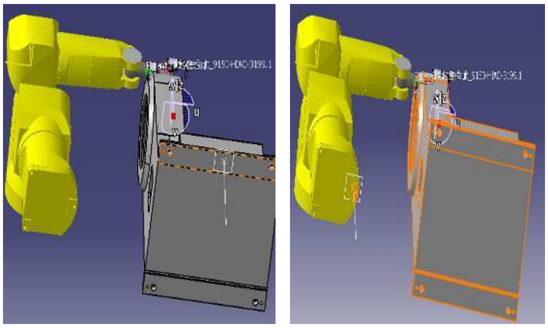


Fig. 18 Illustration of alignment of part and robot in device building

6.4 Adding new task to the robot program

1. Add new task to the robot's program 

Figure 19 illustrates the selection of a new robot task from TaskList under the PPR tree.

- Choose the robot

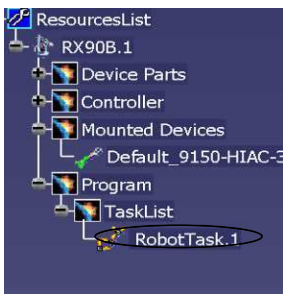


Fig. 19 Illustration of robot task selection under PPR tree.

6.5 Automatic generating tag points

1. Click Create follows Path activity 

- Choose RobotTask.1 from the PPR tree.
- Choose the edges/curves you want to generate tag points automatically.
- Choose the part as product name.
- Press Finish.

Figure 20 illustrates the selection of follow path activity on the part. User can change chord length and can increase or decrease the number of tags generated on the edges.

Fig. 20 Snapshots windows for the selection of follow path activity under robot task

2. Interpolate Tags after changing the orientation.

- Click Robot Teach  then choose robot.
- Choose task "follow path activity".
- Double Click on the follow path activity or click modify.

Figure 21 shows the robot teach window. Users are able to modify the tag points orientation as per requirement.

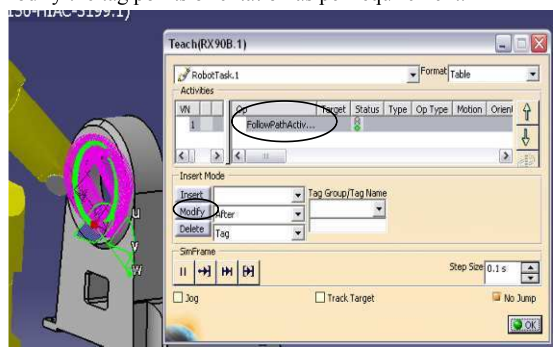


Fig. 21 Snapshots of robot tech window in DELMIA

6.6 Adjusting & interpolating the tag points

- Choose modify from the new window "Teach Continuous Path", as shown in figure 22. In the small window at the right choose first node then adjust the compass/coordinate tool orientation.
- Click set. Repeat same procedure for the end node.
- Press OK.

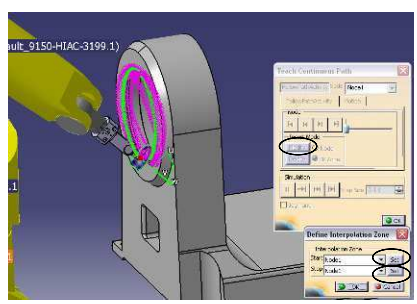


Fig. 22 Snapshots of Teach Continuous Path window.

- After interpolation. The user can find the desired interpolated path as shown in figure 23.

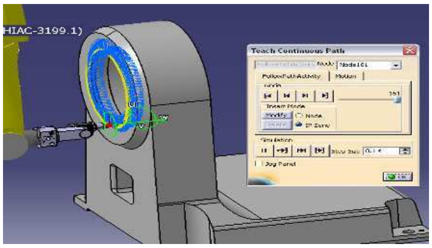


Fig. 23 Illustration of interpolated path after modification

6.7 Simulating for testing the scenario/process.

Figure 24 illustrates simulation of the robot task. It is also possible to check for clash detection during simulation.

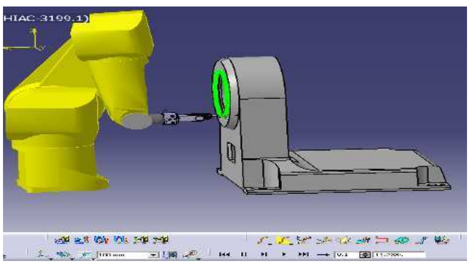


Fig. 24 Snapshot of simulation window for a predefined robot task

Check for Clash detection by 

1. Change the "Singular/Out of Limit Configuration" To Configuration (Good /non Singular).

Figure 25 illustrates teach and jog configuration windows for the robot orientation in DELMIA.

- Click the "jog" Checkbox as shown below and choose Cartesian card.
- Change the configuration to "Good".

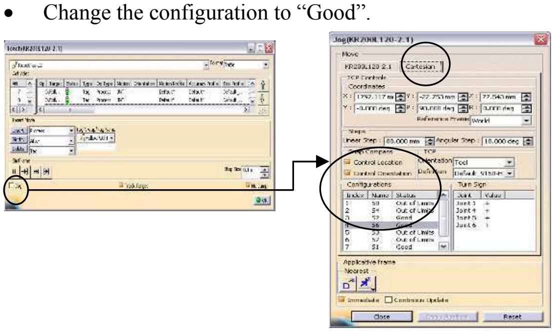


Fig. 25 Snapshots of Teach and Jog configuration windows for robot joint configuration.

2. Change the position of the tool if the user doesn't find any configuration 'Good'.

VII. STEPS FOR PART IN HAND APPROACH USING DELMIA V5R17

7.1. Inserting all the components of the process in Device task definition.
7.2. Adjusting the positions of every component.
7.3. Attaching the tool to the robot.
7.4. Adding new task to the robot program.
7.5. Generating automatic tag points.
7.6. Adjusting & interpolating the tag points.
7.7. Simulating for testing the scenario/process.

Figure 26 illustrates some part in hand configuration screen shots in DELMIA V5 R17.

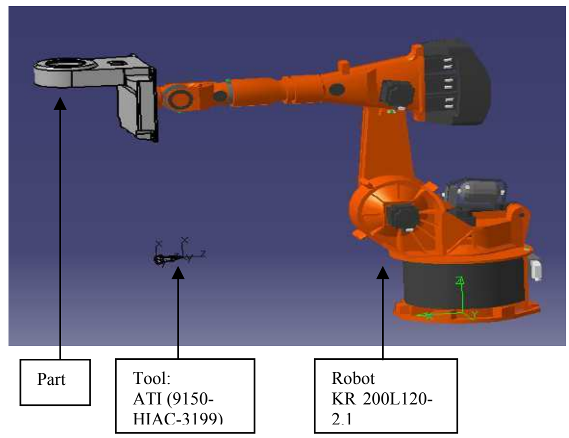


Fig. 26 Screenshots of part in hand configuration mode in DELMIA

A PROCEDURES IN DETAILS

7.1 Inserting all the components of the process in Device task definition.

1. Go to file – open "basesolid.igs".
2. Save this part as basesolid.cat part.
3. Click on Start-Resource – Device Building as shown in figure 27.

Fig. 27 Screenshots for start up of device building

Fig. 28 Screenshots for CATIA product

4. Right click on the P.P.R Tree on Application →Exisiting component → Choose the Part name “basesolid.cat part” as shown in figure 28
5. Save it as basesolid.cat Product.
6. Start-Resource – Device Task Definition.
7. Insert Product. 
8. Choose part name and tool name.
9. Insert Resource. 

7.2 Adjusting the positions of every component.

- Choose the last part of the robot as parent.
- Choose the product/part as a child.
- Press ok.

Figure 29 illustrates the parent child relationship between robot and part.

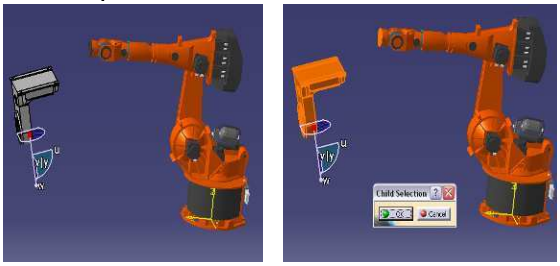
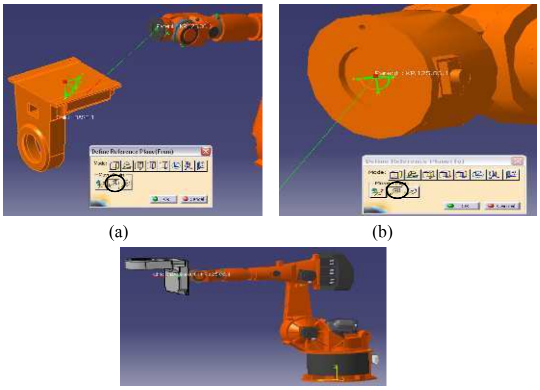


Fig. 29 Screenshot of the parent-child relationship between part and robot.

1. Snap Resource: Attach the part to the end of the robot. 

- Choose the part first.
- Choose icon “Define origin at point or center of face” to set up the coordinate as depicted in figure 30(a).
- Choose the last tip of the robot [not whole robot body] as shown in figure 30(b).
- Choose icon “Define origin at point or center of face” to set up the coordinate again.
- Check that the coordinates coincide as shown in figure 30(c).



(a) (b)

(c)

Fig. 30 Procedures for snapping the part at the end effector of the robot

2. Save the initial State 

7.3 Attaching the tool to the robot.

1. Attach the tool as fixed 

Figure 31 illustrates screenshots of the dress up window.

Fig. 31 Screenshots of the dress up window for the Part in Hand configuration.

7.4 Adding new task to the robot program.

1. Add New Robot Task 

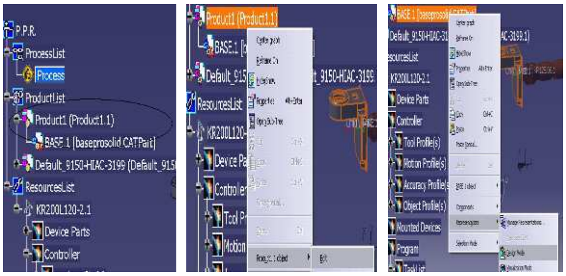
Figure 32 illustrates the robot task under TaskList on PPR tree.

Fig. 32 Screenshots of robot task on PPR tree

2. Change the mode for the part from “Visualization Mode” to “Design Mode”.

Figure 33 shows the steps of mode change from Visualization to Design.

- Right click on Product1 (Product1.1) in PPR tree as shown in figure 33(b).
- From Product1.obj click Edit.
- Then Right click on Base.1 (basesolid.CATpart) —Representation—Design Mode.



(a) (b) (c)

Fig. 33 Screenshots for changing the mode from Visualization to Design.

Find the final shape of the tree/branch of the part in “Design Mode” as shown in figure 34.

Fig. 34 Screenshots for final shape of the part in Design Mode

3. Return to the "Device Task Definition" Environment

- Double Click on Process in PPR tree as shown in figure 35.

Fig. 35 Screenshots of process under PPR tree

7.5 Generating automatic tag points.

1. Click Create follow Path activity 

- Choose the task from PPR tree under program as is illustrated in figure 36.
- Choose the Task from the PPR to add follow path activity to.

Fig. 36 Screenshots of robot task under PPR tree

2. Choose the edge of part to create the path

- Choose any edge for follow path activity as shown in figure 37.

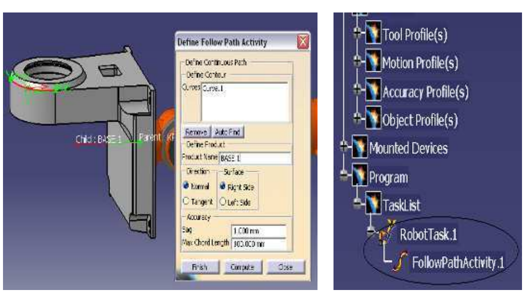


Fig. 37 Screenshots of follow path activity window and its branch under PPR tree

3. Generate new task from follow path activity.

- Right click on Follow Path Activity.
- Follow Path Activity.obj –choose Create New Robot Task as shown in figure 38.

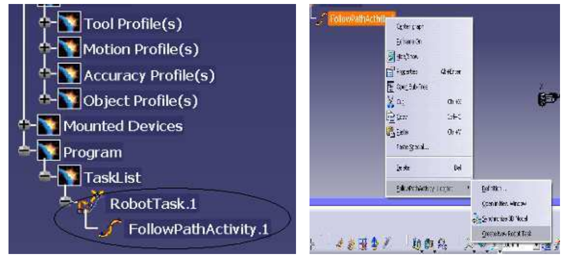


Fig. 38 Screenshots of creation of new robot task from follow path activity

4. Make Parent/child relation between the Tag Group and the part.

- Choose the part as parent.
- Then choose the tag group from PRR tree as child.

7.6 Adjusting & interpolating the tag points

1. Adjust the tag coordinate whether by Tag Transformation or teach a device as per screenshots shown in figure 39.

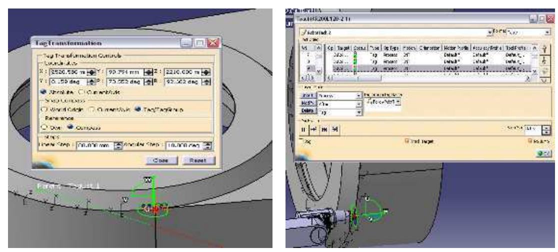


Fig. 39 Screenshots of Tag Transformation and Teach Robot Device.

- In case of discontinuous motion – choose all operation as shown in figure 40.
- Right click choose motion → Linear.

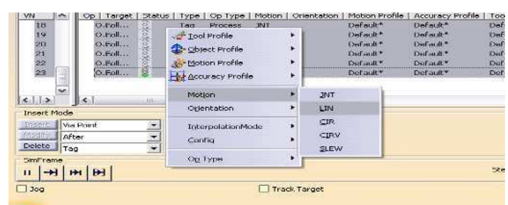


Fig. 40 Screenshots for selection of all tags in linear motion for predefined activity.

2. Interpolate the path 

- Choose the Tag Group

Interpolation of the tag points of predefined path

- Select all tag points and click interpolate as shown in figure 41.

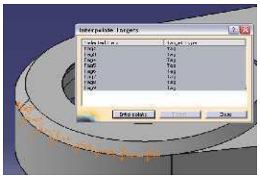


Fig. 41 screenshots windows for interpolation of tag points

3. Add one/two points for pure contact between part and the tool.

7.7 Simulating for testing the scenario/process 

Check the whole task/scenario with Teach Window or With Robot Task simulation as shown below in figure 42.

Figure 42 simulation toolbar

- User can also check for Clash detection by clicking the clash detection icon. 

1. User might change the "Singular/Out of Limit Configuration" to "Configuration Good /non Singular."

- Click "jog" Checkbox as shown below in figure 43 and choose Cartesian card
- Change the configuration to "Good" as shown in figure 43.

2. Change the position of the tool if don't find any configuration 'Good'.

3. User might change the singularity* tolerance if the fourth Joint is from 1 to 10 and the case indicates that current configuration is singular.

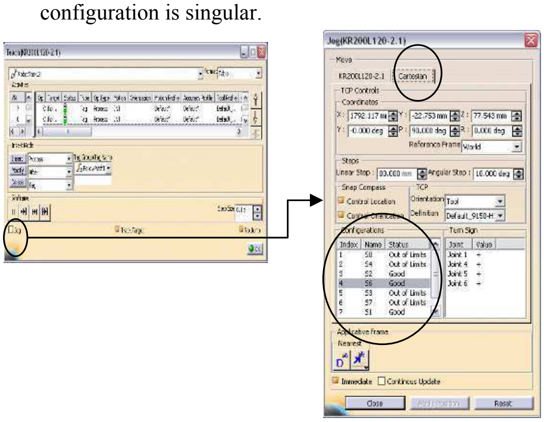


Fig. 43 Screenshots of Teach and Jog configuration window for robot joints configuration

Singularity: DELMIA as a software package is interested in orientation singularity when the fourth joint and sixth joint are on the same axis this is because the fifth joint (The middle joint between them) is at zero distant. It results in a loss of one degree of freedom, as the result of motion whether by the fourth or sixth joint is the same.

VII. Conclusions

DELMIA® enables the offline simulation of tool as well as part in hand deburring approaches in the virtual environment cell. Automated deburring processes eliminate health and safety issues associated with hand held tools. The process is convenient for complex part finishing where it is very hard to reach by conventional machines. Robotics integration provides great flexibility and reliability for the finishing process, because robots allow for more degrees of freedom than other machines. The process has been successfully implemented in several industries.

IX. Future Research

Future work will address surface finishing on the part and generation of the patterns of the points on the surface using micros within DELMIA®. Anticipated work includes establishing prototypes of deburring cells which integrates with DELMIA® in any surface finishing application.

References

[1]. Keith A. Stouffer and Robert Russell, Jr "ADACS: An Advanced Deburring and Chamfering System", prepared at Intelligent Systems Division, National Institute of Standards and Technology, USA.

[2]. Shiakolas, PS., Labalo, D., Fitzgerald, J.M., "RobSurf: A Near Real OLP System for Robotic Surface Finishing", Proceedings of the 7th Mediterranean Conference on Control and Automation (MED99). Haifa, Israel, June, 1999.

[3]. Engel, T. W., Roberts, R. K., Proctor, F. M. "Specification of an Active Force Control Tool for Performing Deburring and Chamfering on a Robot Platform," Proceedings of the International Conference on Industrial Electronics, Control, Instrumentation, and Automation, San Diego, California, November 9-13, 1992.

[4]. Lester E. Godwin, " Programming with Force Control", presented at "The RIA Grinding, Deburring and Finishing workshop St. Paul, Minnesota June 1996.

[5]. Deburring, www.robots.com/applications.php.

[6]. Robotic deburring, www.manufacturingcenter.com.

[7]. Flexiburr, www.ati-ia.com.

[8]. Lawly, T.J., Erlbacher, Edwin A., "Robotic Grinding of Parts to Remove Machine tool Scallops and Mismatch", Maintaining and Supporting an Aircraft Fleet Conference, Dayton, Ohio, June, 1992.

[9]. Automated deburring with the finesse if hand Finishing, www.robotworx.com.

[10] Erlbacher, Edwin A.," A Discussion of Passive And Active Pneumatic Constant Force Devices", session Twenty of the International Robots & Vision Automation Conference, Detroit, Michigan, April, 1993.

[11]. Delmia helps documentation.

Index

M. Iskander et al. (eds.), *Technological Developments in Education and Automation*,
DOI 10.1007/978-90-481-3656-8,

O

P

Q

R

S