AUTOMATIC LANGUAGE IDENTIFICATION

CONTROL AND MEASUREMENT, INDUSTRIAL. SEE INDUSTRIAL MEASUREMENT AND CONTROL. **MEASUREMENT AND CONTROL, INDUSTRIAL. SEE** INDUSTRIAL MEASUREMENT AND CONTROL.

Automatic language identification is the process by which the language of digitized spoken words is recognized by a computer. It is one of several processes in which information is extracted automatically from a speech signal (see SPEECH RECOGNITION; SPEAKER IDENTIFICATION).

Language-identification (LID) applications fall into two main categories: preprocessing for machine systems and preprocessing for human listeners. Figure 1 shows a hotel lobby or international airport of the future that employs a multilingual, voice-controlled retrieval system for travel information. If no mode of input other than speech is used, then the system must be capable of determining the language of the speech commands, either while the system is recognizing the commands or before it has recognized the commands. Determining the language during recognition would require many speech recognizers (one for each language) running in parallel. Because tens or even hundreds of input languages would need to be supported, the cost of the required real-time hardware might prove prohibitive. Instead, a LID system could be used first to list the most likely languages of the speech commands quickly. Then the few most appropriate language-dependent speech-recognition models could be loaded and run. A final LID determination would be made only after speech recognition was complete.

Figure 2 illustrates the second category of LID applications: preprocessing for human listeners. In this case, LID routes an incoming telephone call to a human switchboard operator fluent in the language of the caller. Today, for example, AT&T offers a *Language Line* interpreter service to, among others, police departments handling emergency calls. When a caller to *Language Line* does not speak English, a human operator must attempt to route the call to an appropriate interpreter. Much of the process is trial and error (for example, recordings of greetings in various languages can be used) and can require several connections to find a human interpreter who understands the caller's language. As reported by Muthusamy et al. (1) when callers to *Language Line* do not speak English, the delay in finding a suitable interpreter can be on the order of minutes, which could prove devastating in an emergency. Thus, a LID system that could quickly determine the most likely languages of the caller could reduce the time required to find an appropriate interpreter by one or two orders of magnitude.

LANGUAGE-IDENTIFICATION CUES

Both humans and machines can use a variety of cues to distinguish one language from another. The reader is referred to the linguistics literature (2–4) for in-depth discussions of how languages differ, from one another and to Muthusamy et al. (5), who have measured how well humans can perform language identification. To summarize, languages vary in the following characteristics:

- *Phonology*. A *phoneme* is an underlying mental representation of a phonological unit in a language. For example, the eight phonemes that comprise the word *celebrate* are **/s eh I ix b r ey t/**. A *phone* is a realization of an acoustic–phonetic unit or segment. It is the actual sound produced when a speaker is thinking of speaking a phoneme. The phones that comprise the word *celebrate* might be [**s eh 1 ax bcl b r ey q**]. As documented by linguists, phone and phoneme sets differ from one language to another, even though many languages share a common subset of phones and phonemes. Phone and phoneme frequencies of occurrence may also differ; that is, a phone may occur in two languages, but it may be more frequent in one language than the other. Phonotactics, that is, the rules governing the sequences of allowable phones and phonemes, can also be different.
- *Morphology*. The word roots and lexicons are usually different from language to language. Each language has its own vocabulary and its own manner of forming words.
- *Syntax*. The sentence patterns are different among languages. Even when two languages share a word, for example, the word *bin* in English and German, the words that may precede and follow the shared word will be different.
- *Prosody*. Duration of phones and syllables, pitch contours, and stress patterns are different from one language to another.

LANGUAGE IDENTIFICATION SYSTEMS

Research in automatic language identification from speech began in the 1970s. A few representative LID systems are described below. The reader will find references to other LID systems in reviews by Mumusamy et al. (1) and Zissman (6).

Figure 3 shows the two phases of LID. During the training phase, the typical system is presented with examples of speech from a variety of languages. Some systems require only the digitized speech utterances and the corresponding true identities of the languages being spoken. More complicated LID systems may require labeling, that is, either 1) a phonetic transcription (sequence of symbols representing the sounds spoken) 2) an orthographic transcription (the text of the words spoken) along with a phonemic transcription dictionary (mapping of words to prototypical pronunciation) for each training utterance. Producing these transcriptions and dictionaries is an expensive, time-consuming process that usually requires a skilled linguist fluent in the language of interest. Each training speech utterance is converted into a stream of feature vectors. These feature vectors are computed from short windows of the speech waveform (e.g., 20 ms) during which the speech signal is assumed to be somewhat stationary. The feature vectors are recomputed regularly (e.g., every 10 ms) and

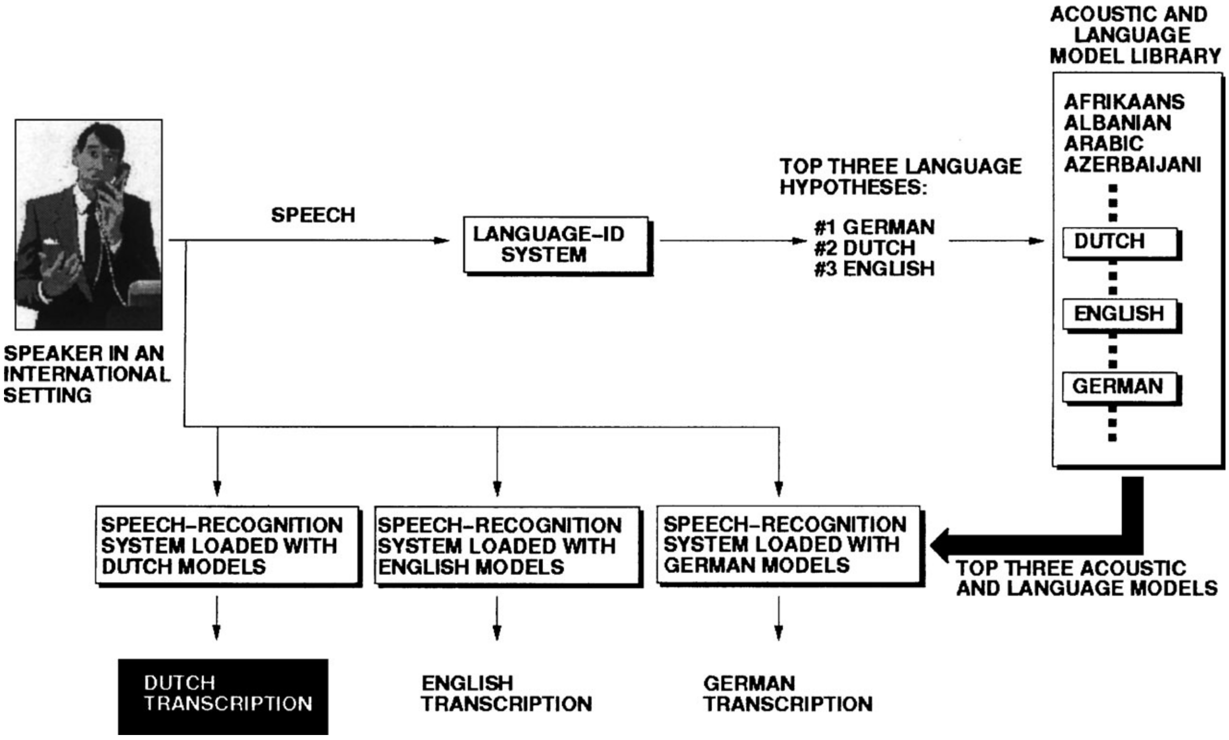
ACOUSTIC AND LANGUAGE MODEL LIBRARY
AFRIKAANS
ALBANIAN
ARABIC
AZERBAIJANI
DUTCH
ENGLISH
GERMAN
SPEECH
LANGUAGE-ID SYSTEM
TOP THREE LANGUAGE HYPOTHESES:
#1 GERMAN
#2 DUTCH
#3 ENGLISH
SPEAKER IN AN INTERNATIONAL SETTING
SPEECH-RECOGNITION SYSTEM LOADED WITH DUTCH MODELS
SPEECH-RECOGNITION SYSTEM LOADED WITH ENGLISH MODELS
SPEECH-RECOGNITION SYSTEM LOADED WITH GERMAN MODELS
TOP THREE ACOUSTIC AND LANGUAGE MODELS
DUTCH TRANSCRIPTION
ENGLISH TRANSCRIPTION
GERMAN TRANSCRIPTION

Figure 1. A LID system as a front end to a set of real-time speech recognizers. The LID system outputs its three best guesses of the language of the spoken message (in this case, German, Dutch, and English). Speech recognizers are loaded with models for these three languages and make the final LID decision (in this case, Dutch) after decoding the speech utterance.

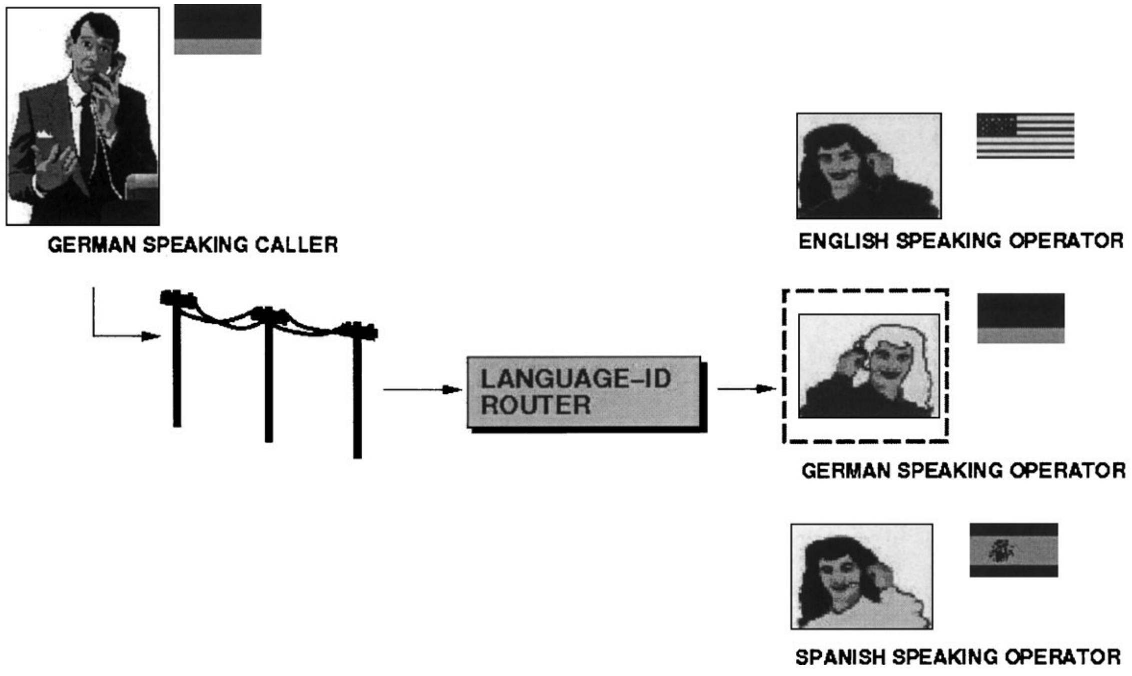

Figure 2. A LID system as a front end to a multilingual group of directory-assistance or emergency operators. The LID system routes an incoming call to a switchboard operator fluent in the corresponding language.

contain spectral or cepstral information about the speech signal (the cepstrum is the inverse Fourier transform of the log magnitude spectrum; it is used in many speech processing applications). The training algorithm analyzes a sequence of such vectors and produces one or more models for each language. These models are intended to represent a set of fundamental characteristics for each language of the training speech. The sets are used during the next phase of the LID process.

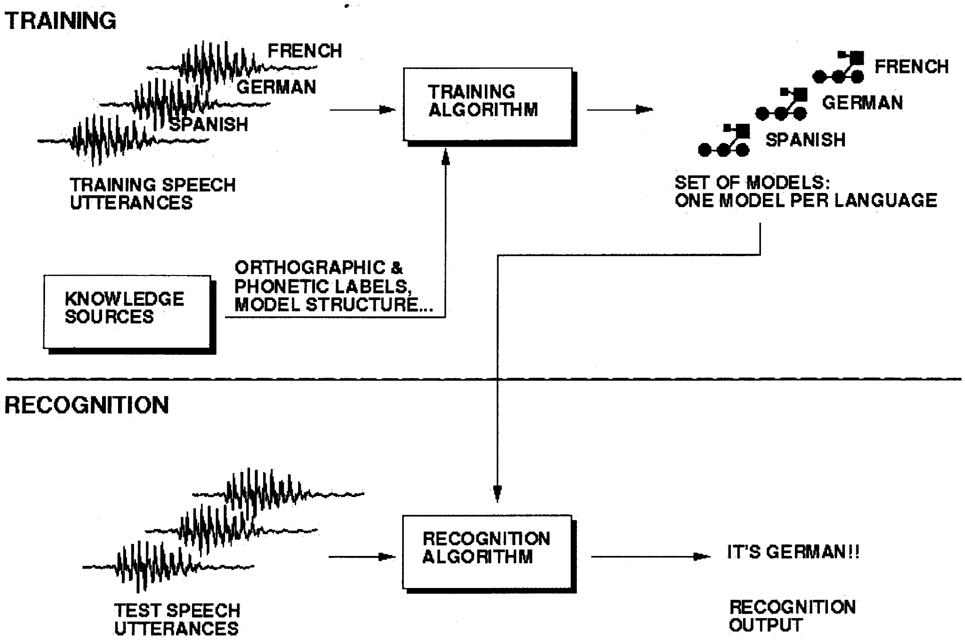

Figure 3. The two phases of language identification. During training, speech waveforms are analyzed and language-dependent models are produced. During recognition, a new speech utterance is processed and compared with the models produced during training. The language of the speech utterance is hypothesized.

During the recognition phase of LID, feature vectors computed from a new utterance are compared with the models for each language. The likelihood that the new utterance was spoken in the same language as the speech used to train each model is computed and the most likely model is found. The language of the new example is hypothesized to be the same as the language of the most likely model.

Spectral-Similarity Approaches

In the earliest automatic LID systems, developers capitalized on the differences in spectral content among languages, exploiting the fact that speech spoken in different languages contains different phonemes and phones. To train these systems, a set of prototypical short-term spectra was computed and extracted from training speech utterances. During recognition, test speech spectra were computed and compared with the training prototypes. The language of the test speech was hypothesized as the language having training spectra that best matched the test spectra.

Several variations on this spectral-similarity theme existed. The training and testing spectra could be used directly as feature vectors, or they could be used instead to compute formant-based or cepstral feature vectors. The training exemplars could be chosen either directly from the training speech or could be synthesized through the use of K-means clustering. The spectral similarity could be calculated by the Euclidean, Mahalanobis, or other distance metric. Examples of spectral similarity LID systems have been proposed and developed by Cimarusti and Ives (7) Foil (8), Goodman et al. (9), and Sugiyama (10).

To compute the similarity between a test utterance and a training model, most early spectral-similarity systems calculated the distance between each test utterance vector and each training exemplar. The distance between each test vector and its closest exemplar was accumulated as an overall distance, and the language model having the lowest overall distance was found. In a generalization of this vector quantization approach to LID, Riek et al. (11), Nakagawa et al. (12), and Zissman (13) applied Gaussian mixture classifiers to language identification. They assumed each feature vector is drawn randomly according to a probability density that is a weighted sum of multivariate Gaussian densities. During training, a Gaussian mixture model for the spectral or cepstral feature vectors is created for each language. During recognition, the likelihood of the test utterance feature vectors is computed for each training model. The language of the model having maximum likelihood is hypothesized. The Gaussian mixture approach is a "*soft*" vector-quantization, where more than one exemplar created during training impacts the scoring of each test vector.

Whereas the language identification systems described above perform primarily static classification, hidden Markov models (HMMs) (14) which can model sequential characteristics of speech production, have also been applied to LID. HMM-based language identification was first proposed by House and Neuburg (15). Savic et al. (16), Riek et al. (11), Nakagawa et al. (12), and Zissman (13) all applied HMMs to spectral and cepstral feature vectors. In these systems, HMM training was performed on unlabeled training speech (i.e., training speech with no corresponding phonetic or phonemic transcription). Riek et al. and Zissman found that HMM systems trained in this unsupervised manner (i.e., with unlabeled speech) did not perform as well as some of the static classifiers that had been tested, although Nakagawa et al. eventually obtained better performance using HMMs (17).

Li (18) has proposed using novel features for spectral-similarity LID. In his system, the syllable nuclei (i.e., vowels and syllabic consonants) for each speech utterance are located automatically and feature vectors are computed near the spectral nuclei for each training speaker. During testing, syllable nuclei of the test utterance are located and feature vectors are extracted. The set of feature vectors for each training speaker is compared with the feature vectors of the test speech, and the training speaker having the most similar set of feature vectors is found. The language used by the speaker of that set of training vectors is hypothesized as the language of the test utterance.

Recently, Torres-Carrasquillo et al. (19) and Kohler and Kennedy (20) have proposed a Gaussian mixture model approach that incorporates additional information about the speech dynamics. By stacking delta-cepstral vectors in each feature vector, a process known as shifted-delta cepstra [SDC], and increasing the mixture model order, this approach tries to overcome some problems with static classification in previous approaches. Burget et al. (21) has obtained even better performance using a discriminative training approach.

Campbell et al. (22) has incorporated the SDC feature processing technique into a support vector machine classifier. In this work, Campbell et al. generate a feature vector for each utterance of interest using a degree 3 monomial expansion. Each feature vector for each language of interest is then used in a “one vs. all” training strategy. For example, in the case of English, all utterances for English are used for class A, whereas all other utterances for all competing languages are pooled into class B. The resulting model is used as the English model, and the process is repeated for each language of interest.

Phone-Recognition Approaches

Given that different languages have different phone inventories, many researchers have built LID systems that hypothesize exactly which phones are being spoken as a function of time and determine the language based on the statistics of that phone sequence. For example, Lamel and Gauvain built two HMM-based phone recognizers: one in English and another in French (23). These phone recognizers were then run over test data spoken either in English or French. Lamel and Gauvain found that the likelihood scores emanating from language-dependent phone recognizers can be used to discriminate between English and French speech. Muthusamy et al. ran a similar system on English and Japanese spontaneous telephone speech (24).

The novelty of these phone-based systems was the incorporation of more knowledge into the LID system. Both Lamel et al. and Muthusamy et al. trained their systems with multilahguage, phonetically labeled corpora. Because the systems require phonetically labeled training speech utterances in each language, as compared with the spectral-similarity systems that do not require such labels, it can be more difficult to incorporate new languages into the language-recognition process.

To make phone-recognition–based LID systems easier to train, one can use a single-language phone recognizer as a front end to a system that uses phonotactic scores to perform LID. Phonotactics are the language-dependent set of constraints specifying which phonemes are allowed to follow other phonemes. For example, the German word *spiel*, which is pronounced **/sh p iy l/** and might be spelled in English as *shpeel*, begins with a consonant cluster **/sh p/** that cannot occur in English (except if one syllable ends in **/sh/** and the next begins with **/p/**, or in a compound word like *flashpoint*). This approach is similar to that used by D’Amore and Mah (25), Kimbrell (26), Schmitt (27), and Damashek (28), who have used n-gram analysis of text documents to perform language and topic identification and clustering. By “tokenizing” the speech message, that is, converting the input waveform to a sequence of phone symbols, the statistics of the resulting symbol sequences can be used to perform language identification. Figure 4 shows the systems of Hazen and Zue (29) and Zissman and Singer (30), who each developed LID systems that use one single-language front-end phone recognizer. An important finding of these researchers was that LID could be performed successfully even when the front-end phone recognizer was not trained on speech spoken in the languages to be recognized. For example, accurate Spanish versus Japanese LID can be performed using only an English phone recognizer. Zissman and Singer (30) and Yan and Barnard(31) have extended this work to systems containing multiple single-language front ends, where there need not be a front end in each language to be identified. Figure 5 shows an example of these types of systems. Meanwhile, Hazen and Zue (32) and Navratil and Zuhlke (33) have pursued LID systems that use a single multilanguage front-end phone recognizer.

In the last few years, the work of Zissman and Singer (30) has been extended by Gauvain et al. (34), incorporating a more general approach at the phone-recognizer stage. Instead of using the best phone sequences, that is, the most likely, set of phones for the given utterance, Gauvain et al. use lattices, allowing for a more general decoding of the incoming speech. The work by Gauvain et al. show better performance on similar data sets when compared with Zissman and Singer’s system.

Speech-to-Text Approaches

By adding even more knowledge to the system, researchers hope to obtain even better LID performance. Mendoza et al. (35), Schultz et al. (36), and Hieronymus and Kadambe (37) have shown that speech-to-text (STT) systems can be used for LID. During training, one speech recognizer for each language is created. During testing, each of these recognizers operates in parallel. The one yielding output with highest likelihood is selected as the winning recognizer—the language used to train that recognizer is the hypothesized language of the utterance. Such systems hold the promise of high-quality,language identification because they use higher level knowledge (words and word sequences) rather than lower level knowledge (phones and phone sequences) to make the LID decision. Furthermore, one obtains a transcription of the utterance as a byproduct of LID. However, these systems require many hours of labeled training data in each target language and are the most computationally complex of the algorithms proposed.

Control and measurement, industrial. See INDUSTRIAL MEASUREMENT AND CONTROL. **Measurement and control, industrial. See** INDUSTRIAL MEASUREMENT AND CONTRC

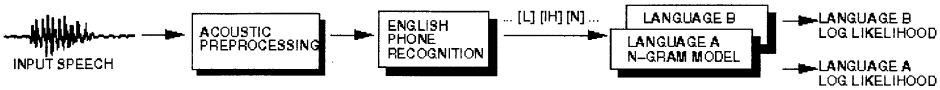

Figure 4. The phone recognition followed by phone frequency and phone sequence language modeling LID system. Phone recognition is performed in one language, in this case, English. Phone frequency and sequence statistics are used to determine the language of the speech utterance.

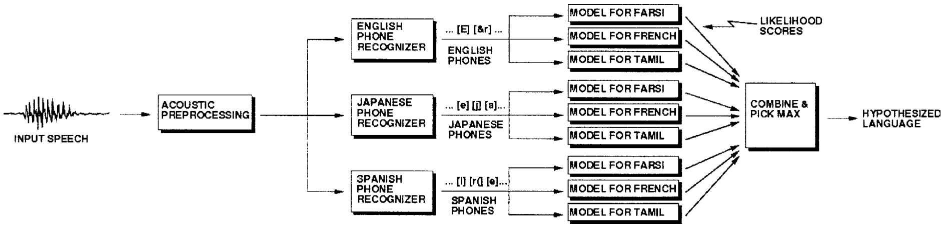

Figure 5. A LID system that uses several phone recognizers in parallel.

EVALUATIONS

Since 1993, the National Institute of Standards and Technology (NIST) of the U.S. Department of Commerce has sponsored formal evaluations of LID systems. At first, these evaluations were conducted using the Oregon Graduate Institute Multi-Language Telephone Speech (OGI-TS) Corpus (38). The OGI-TS corpus contains 90 speech messages in each of the following 11 languages: English, Farsi, French, German, Hindi, Japanese, Korean, Mandarin, Spanish, Tamil, and Vietnamese. Each message is spoken by a unique speaker and comprises responses to 10 prompts. For NIST evaluations, the monologue speech evoked by the prompt "Speak about any topic of your choice" is used for both training and testing. No speaker speaks more than one message or more than one language, and each speaker's message was spoken over a unique long-distance telephone channel. Phonetically transcribed training data are available for six OGI languages (English, German, Hindi, Japanese, Mandarin, and Spanish).

Performance of the best systems from the 1993, 1994, and 1995 NIST evaluations is shown in Fig. 6. This performance represents each system's first pass over the evaluation data, which means that no system-tuning to the evaluation data was possible. For utterances having a duration of either 45 s or 10 s, the best systems can discriminate between two languages with 4% and 2% error, respectively. This error rate is the average computed over all language pairs with English, for example, English versus Farsi, English versus French, and so on. When tested on nine-language forced-choice classification, error rates of 12% and 23% have been obtained on 45 s and 10 s utterances, respectively. The syllabic-feature system developed by Li and the systems with multiple phone recognizers followed by phonotactic language modeling developed by Zissman and Yan have exhibited the best performance in these evaluations. Error rate has decreased over time, which indicates that research has improved system performance.

Starting in 1996, the NIST evaluations have employed the Linguistic Data Consortium's CALLFRIEND corpus. CALLFRIEND comprises two-speaker, unprompted, conversational speech messages between friends. North American long-distance telephone conversations were recorded in each of 12 languages (the same 11 languages as OGI-TS plus Arabic). No speaker occurs in more than one conversation. In the 1996 evaluation, the multiple phone recognizer followed by language modeling systems of Yan and Zissman performed best. The error rates on 30 s and 10 s utterances were 5% and 13% for pairwise classification. These same systems obtained 23% and 46% error rates for 12-language classification. The higher error rates on CALLFRIEND are from the informal conversational style of CALLFRIEND versus the more formal monologue style of OGI-TS.

After the 1996 evaluation, NIST evaluations were not conducted until 2003. In the 2003 evaluation, the CALLFRIEND corpus was used again by including an additional set of conversations not previously exposed during the 1996 evaluation. Two new trends emerged from the 2003 evaluation: 1) Spectral similarity approaches, particularly Gaussian mixture models and support vector machines, were proven to provide competitive performance to the phone-recognition based approaches; and 2) system combination, also known as system fusion, rather than individual standalone systems, were shown to provide additional performance over the individual constituents. The system combination concept arises from the fact that errors observed within the individual systems can be corrected as long as they occur independently. An example of the results obtained by Singer et al. (39) for the 2003 evaluation set is shown in Fig. 7.

The STT-based LID systems have not been fully evaluated at NIST evaluations, because orthographically and phonetically labeled speech corpora have not been available in each requisite language. However, preliminary results on selected language pairs of the OGI-TS corpus indicate near-perfect performance. As labeled corpora become available in more languages, implementation and evaluation of STT-based LID systems will become more feasible. Whether the performance they will afford will be worth their computational complexity remains to be seen.

CONCLUSIONS

Since the 1970s, language identification systems have become more accurate and more complex. Systems can per-

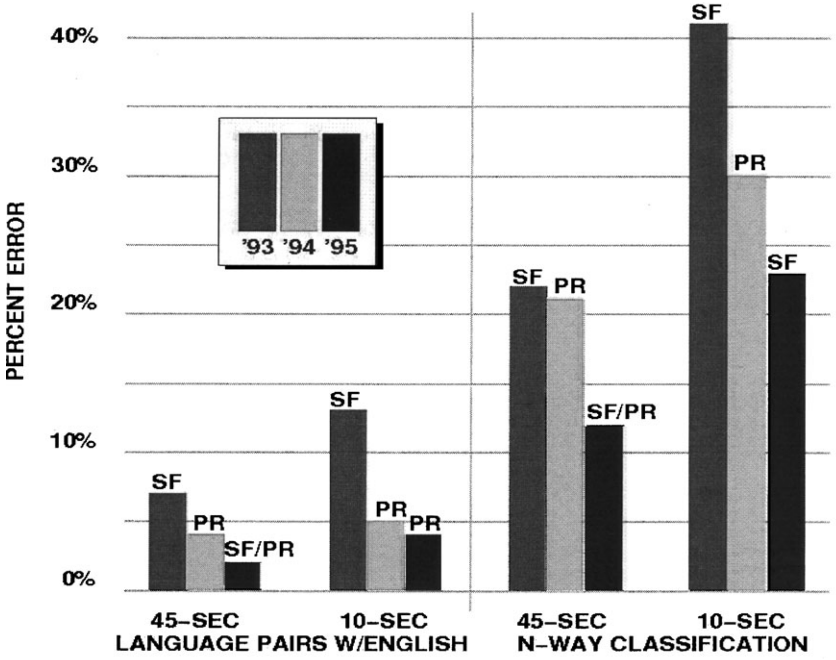

Figure 6. Error rates of the best LID systems at three NIST evaluations. Performance is shown on the left for average two-alternative, forced-choice classification of the various OGI-TS languages with English. "N-way" classification refers to 10-alternative, forced-choice performance in 1993; 11-alternative, forced-choice performance in 1994; and 9-alternative, forced-choice performance in 1995. "SF" indicates a syllabic feature system. "PR" indicates phone recognition followed by a language modeling system.

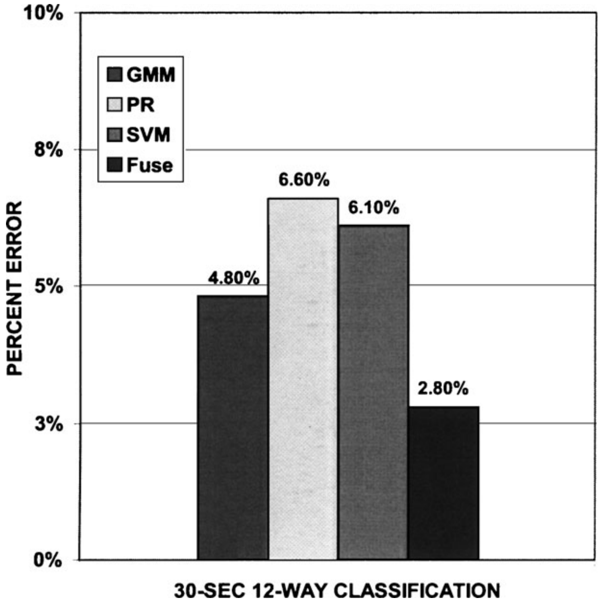

Figure 7. Error rates of the best LID system at the NIST 2003 evaluation. Performance is shown for all individual components of the system and for the combination of the three systems, on a 12-alter native, forced-choice scenario. "GMM" indicates Gaussian mixture models. "PR" indicates phone recognition followed by the language modeling system. "SVM" indicates a support vector machine. "Fuse" indicates the combination of the previous three systems.

form two-alternative, forced-choice identification on extemporaneous monologue almost perfectly, with the newest systems performing 12-way identification with roughly 3% error. As shown from evaluations in 2003, error rates on conversational speech have been reduced compared with 1996.

Although initially the improved performance of LID systems was from their use of higher levels of linguistic information, in recent years, systems that do not require high-level information have been steadily improving. Recent results in the 2003 evaluation show the spectral-similarity systems outperforming phone-recognition approaches. Additionally, the spectral-similarity approaches seem to be complementary to the phone-recognition approaches, as shown by the improved performance obtained by the combination of the systems.

Still, as the number of potential applications grows, faster implementations are needed along with systems that can easily be adapted to new conditions and languages.

ACKNOWLEDGMENT

This work is sponsored by the Department of Defense under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

BIBLIOGRAPHY

1. Muthusamy, Y. K.; Barnard, E.; Cole, R. A.; Reviewing Automatic Language Identification. *IEEE Signal Process. Mag.* 1994, **11**(4),pp 33–41.
2. Comprie, B. *The World's Major Languages*; Oxford University Press: New York, 1990.
3. Crystal, D. *The Cambridge Encyclopedia of Language*; Cambridge University Press: Cambridge, UK, 1987.
4. Fromkin, V.; Rodman, R. *An Introduction to Language*; Harcourt Brace Jovanovich: Orlando, FL, 1993.
5. Muthusamy, Y. K.; Jain, N.; Cole, R. A. Perceptual Benchmarks for Automatic Language Identification; *ICASSP 1994 Proc:*, 333–336.
6. Zissman, M. A. Comparison of Four Approaches to Automatic Language Identification of Telephone Speech. *IEEE Trans. Speech Audio Proc.*, 1996, **4**(1), 31–44.
7. Cimarusti, D.; Ives, R. B. Development of an Automatic Identification System of Spoken Languages: Phase I; *ICASSP 1982 Proc.*;pp 1661–1663.
8. Foil, J. T. Language Identification Using Noisy Speech; *ICASSP 1986 Proc.*; **2**, 861–864.
9. Goodman, F. J.; Martin, A. F.; Wohlford, R. E. Improved Automatic Language Identification in Noisy Speech.; *ICASSP 1989 Prco.*; **1**,pp 528–531.
10. Sugiyama, M. Automatic Language Recognition using Acoustic Features; *ICASSP 1991 Proc.*; **2**,pp 813–816.
11. Riek, L.; Mistretta, W.; Morgan, D. Experiments in Language Identification.Technical Report SPCOT-91-002, Lockheed Sanders, Inc., Nashua, NH, December 1991.
12. Nakagawa, S.; Ueda, Y.; Seino, T. Speaker-Independent, Text-Independent Language Identification by HMM; *ICSLP 1992 Proc.*; **2**,pp 1011–1014.
13. Zissman, M. A. Automatic Language Identification using Gaussian Mixture and Hidden Markov Models; *ICASSP 1993 Proc.*; **2**,pp 399–402.
14. Rabiner, L. R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE*; 1989, **7**(2),pp 257–286.
15. Hoouse, A. S.; Neuburg, E. P. Toward Automatic Identification of the Language of an Utterance. I. Preliminary Methodological Considerations; *J. Acoust. Soc. Amer.*; 1977, **62**(3),pp 705–730.
16. Savic, M.; Acosta, E.; Gupta, S. K. An Automatic Language Identification System; *ICASSP 1991 Proc.*, **2**,pp 817–820.
17. Nakagawa, S.; Seino, T.; Ueda, Y. Spoken Language Identification by Ergodic HMMs and its State Sequences; *Electron. Commun. J. Part 3*, 1994, **7**(6), 70–79.
18. Li, K.-P. Automatic Language Identification using Syllabic Spectral Feature; *ICASSP 1994 Proc.*, **1**, 297–300.
19. Torres-Carrasquillo, P. A.; Singer, E.; Kohler, M. A.; Greene, R. J.; Reynolds, D. A.; Deller, Jr., J. R. Approaches to Language Identification using Gaussian Mixture Models and Shifted Delta Cepstral Features, *ICSLP 2002 Proc.*;pp 33–36.
20. Kohler, M. A.; Kennedy, M. Language Identification using Shifted Delta Cepstra.; *MWSCAS-2002 Proc*; **3**,pp 69–72.
21. Burget, L.; Matejka, P.; Cernocky, J. Discriminative Training Techniques for Acoustic Language Identification. *ICASSP 2006 Proc.*;pp 209–212.
22. Campbell, W. M.; Singer, E.; Torres-Carrasquillo, P. A.; Reynolds, D. A. Language Recognition with Support Vector Machines.; *Proc. Odyssey 2004: The Speaker and Language Recognition Workshop*;pp 41–44.
23. Lamel, L. F.; Gauvain, J.-L. Cross-Lingual Experiments with Phone Recognition. *ICASSP '93 Proc.*;pp 507–510.
24. Muthusamy, Y. et al., A Comparison of Approaches to Automatic Language Identification using Telephone Speech., *Eurospeech 1993 Proceedings*, **2**,pp 1307–1310.
25. D'Amore, R. J.; Mah, C. P. One-Time Complete Indexing of Text: Theory and Practice; *Proc. of the Eighth Intl. ACN Conf. on Res. and Dev. in Information Retrieval*; 1985, pp 155–164.
26. Kimbrell, R. E. Searching for Text? Send an n-gram! *Byte*, 1988, **13**(5),pp 297–312.
27. Schmitt, J. C.Trigram-Based Method of Language Identification.U.S. Patent 5,062,143,October 1991.
28. Damashek, M. Gauging Similarity with n-grams: Language-Independent Categorization of Text *Science*, 1995, **267**,pp 843–848.
29. Hazen, T. J.; Zue, V. W. Automatic Language Identification using a Segment-Based Approach; *Eurospeech 1993 Proc.*; **2**,pp 1303–1306.
30. Zissman, M. A.; Singer, E. Automatic Language Identification of Telephone Speech Messages using Phoneme Recognition and n-Gram Modeling.; *ICASSP 1994 Proc.*; **1**pp 305–308.
31. Yan, Y.; Barnard, E. An Approach to Automatic Language Identification based on Language-Dependent Phone Recognition; *ICASSP 1995 Proc.*; **5**,pp 3511–3514.
32. Hazen, T. J.; Zue, V. W. Recent Improvements in an Approach to Segment-Based Automatic Language Identification; *ICSLP 1994 Proc.*; **4**pp 1883–1886.
33. Navratil, J.; Zuhlke, W. Double Bigram-Decoding in Phonotactic Language Identification; *ICASSP 1997 Proc.*; **2**,pp 1115–1118.
34. Gauvain, J. L.; Messaoudi, A.; Schwenk, H. Language Recognition using Phone Lattices; *ICSLP 2004 Proc.*;pp 1283–1286.
35. Mendoza, S. et al., Automatic Language Identification using Large Vocabulary Continuous Speech Recognition; *ICASSP 1996 Proc.*; **2**,pp 785–788.
36. Schultz, T.; Rogina, I.; Waibel, A., LVCSR-Based Language Identification; *ICASSP 1996 Proc.*; **2**,pp 781–784.
37. Hieronymus, J. L.; Kadambe, S. Robust Spoken Language Identification using Large Vocabulary Speech Recognition; *ICASSP 1997 Proc.*; **2**, 111–114.
38. Muthusamy, Y. K.; Cole, R. A.; Oshika, B. T. The OGI Multi-Language Telephone Speech Corpus; *ICSLP 1992 Proc.*; **2**,pp 895–898.

39. Singer, E.; Torres-Carrasquillo, P. A.; Gleason, T. P.; Campbell, W. M.; Reynolds, D. A. Acoustic, Phonetic, and Discriminative Approaches to Automatic Language Recognition. *Eurospeech 2003 Proc.*;pp 1345–1348.

PEDRO A. TORRES-CARRASQUILLO

MARC A. ZISSMAN
MIT Lincoln Laboratory