Exploratory Analysis and Data Modeling in Functional Neuroimaging

edited by **FRIEDRICH T. SOMMER** and **ANDRZEJ WICHERT**
foreword by **Manfred Spitzer**

Exploratory Analysis and Data Modeling in Functional Neuroimaging

Neural Information Processing Series
Michael I. Jordan, Sara A. Solla, and Terrence J. Sejnowski, editors

Advances in Large Margin Classifiers
Alexander J. Smola, Peter L. Bartlett, Bernhard Schölkopf and Dale Schuurmans, eds., 2000

Advanced Mean Field Methods: Theory and Practice
Manfred Opper and David Saad, eds., 2001

Probabilistic Models of the Brain: Perception and Neural Function
Rajesh P. N. Rao, Bruno A. Olshausen, and Michael S. Lewicki, eds., 2002

Exploratory Analysis and Data Modeling in Functional Neuroimaging
Friedrich T. Sommer and Andrzej Wichert, eds., 2003

Exploratory Analysis and Data Modeling in Functional Neuroimaging

edited by
Friedrich T. Sommer
Andrzej Wichert

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

This book was set in Times Roman by the author using the LaTeX document preparation system.

Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Exploratory analysis and data modeling in functional neuroimaging/ by Friedrich
T. Sommer, Andrzej Wichert.
p. cm. — (Neural information processing series)
Includes bibliographical references and index.
ISBN 0-262-19481-3 (hc : alk. paper)
1. Nervous system—Imaging—Data processing. 2. Nervous system—Imaging—Mathematics.
I. Sommer, Friedrich T. II. Wichert, Andrzej. III. Series
QP357.5.E96 2003
612.8′028-dc21

2002023002

Contents

Series Foreword

The yearly Neural Information Processing Systems (NIPS) workshops bring together scientists with broadly varying backgrounds in statistics, mathematics, computer science, physics, electrical engineering, neuroscience, and cognitive science, unified by a common desire to develop novel computational and statistical strategies for information processing, and to understand the mechanisms for information processing in the brain. As opposed to conferences, these workshops maintain a flexible format that both allows and encourages the presentation and discussion of work in progress, and thus serve as an incubator for the development of important new ideas in this rapidly evolving field.

The Series Editors, in consultation with workshop organizers and members of the NIPS Foundation Board, select specific workshop topics on the basis of scientific excellence, intellectual breadth, and technical impact. Collections of papers chosen and edited by the organizers of specific workshops are built around pedagogical introductory chapters, while research monographs provide comprehensive descriptions of workshop-related topics, to create a series of books that provides a timely, authorative account of the latest developments in the exciting field of neural computation.

Michael I. Jordan, Sara A. Solla, and Terrence J. Sejnowski

This page intentionally left blank

Foreword

When I was searching medical databases about ten years ago for papers on neural networks, I was surprised that more than half of the hits I got with the key words "neural networks" in MEDLINE were about their application as special devices for pattern recognition. Only a small part of the hits concerned the brain and models of brain function. I had expected the opposite to be the case and actually had not thought much about the application of neural networks in medicine. In my view, neural networks are perfect models for understanding the working principles of the brain. They can describe brain function on an abstract, systems neuroscience level, while at the same time they reflect organizational principles of the neuronal substrate (rather than boxes of logical operators or functions).

On both accounts, as models for top-down approaches to understanding brain function and as tools for bottom-up approaches for analysis and data mining of medical data, neural networks made rapid progress: In 1998, the U. S. Food and Drug Administration approved a neural network based device for the screening of gynecological smears for cancerous cells. Intensive care units use neural networks to allocate patients to resources in the most rational way (with cultural differences regarding the interpretation: In the United Kingdom, doctors say they consider the advice from the computer but decide for themselves, whereas in the United States doctors say they usually let the computer decide). Even psychiatrists use neural networks to group symptoms into syndromes. In almost every field of medicine, model based top-down strategies and bottom-up strategies of data mining complement each other. For modeling brain function, back-propagation networks with biologically less-plausible features have given way to more plausible architectures, from Kohonen-maps and adaptive resonance networks to spiking neuron networks. Various models have been used to account for a wide range of psychological phenomena, from simple perception, via reading and attention, to language aquisition, schizophrenia, and autism (cf. Stein and Ludik, 1998; Spitzer, 1999). The impact of these developments on systems neuroscience research has been profound. This book is perhaps the first comprehensive volume summarizing the different roles of neural networks in a new important field of physiology, functional neuroimaging. It describes how neural networks—in a broader sense— act as tools for sophisticated exploratory data analysis on the one hand, and for brain models on the other.

The advent of recent functional neuroimaging techniques has changed the situation in physiology entirely. Physiologists used to be proud of the fact that they did not need any statistics to prove their point with their rather simple measurements.

"You either see an effect or you don't—why run t-tests?" was a common frame of mind within the field. Psychologists were looked down upon, as they had to use statistics, presumably because they used bad methods that produced noisy data. However, functional magnetic resonance imaging (fMRI), high resolution magnetoencephalography (MEG), and event related potentials (ERP) generate signals that are notoriously noisy, and the effect sizes are small. In fact, these methods were developed to rely upon averaging data across time and/or space, and they became practical only when computing power to do the necessary number crunching became available. With refinement of the techniques, they became less reliant on averaging, but only because more sophisticated methods of data analysis were developed, requiring even more computing power than simple averaging.

Finding patterns among noisy signals is a task which the human brain is particularly good at (cf. Huettel et al., 2002), up to the point where it is running the risk of generating superstitious beliefs or even outright delusions. The history of science is full of such apparent patterns, from the channels on Mars "observed" by the astronomer Lowell in 1906 to the many psychosocial "theories" of various scope based upon unwarranted inferences from spurious p-values generated by the analysis of questionnaires. It is therefore equally as important to try to use the brain's strategies for data analysis as it is to reflect upon them and scrutinize what they are able to do and what they are unable to achieve. The more a system—be it the brain or a data mining program—knows about the structure to be discovered, the better the chances of detection. This is why experts in any field, from chess players to paleontologists, are able to recognize salient structures and "see" the imminent attack and the million-year-old tooth in the rubble. This is why we have to turn to models that incorporate what we already know about the brain, in order to extract most of the signal generated under specific conditions. And just as it is easier to see something when you hear a noise coming from the same spot in space, the combination of data from two or more techniques can produce vistas upon the brain that any single method cannot provide. Of course, such multimodal functional neuroimaging techniques require ingenious ways of analyzing data, such that one method can be used to constrain data generated by the other method, and vice versa.

The interpretation of functional neuroimaging data is not unlike the hermeneutic interpretation of a text. You do not see the details if you do not have the "big picture," but you cannot see the big picture without seeing the details. As we know from hermeneutics, the solution to this apparent paradox consists of iterations (what used to be called the hermeneutic circle), i.e. progressive analytic/synthetic steps, whereby each has to be tightly controlled by the data as well as guided by the increasingly clear view of the result. As there are thousands of ways to analyze the complex data generated in systems neuroscience research, models have to be used to guide this process from the very beginning on.

The promise of functional neuroimaging is not less than to tackle one of the most fundamental open questions of understanding the brain, namely how microscopic and macroscopic organization in the brain relate and interact in order to produce

brain function. We may compare the road ahead of brain research with the voyage to the moon. Functional brain imaging then may correspond to an important element of rocket technology. To achieve the entire trip, however, it is necessary to put together many pieces of evidence collected by researchers working in different fields. Blood-flow based techniques alone cannot assess the temporal structure of brain activity, and even bundling all available techniques of brain research, uncovering the neural basis of behavior remains an underconstrained problem.

This book provides an overview of the various contributions theorists can and must deliver in order to apply functional imaging successfully for solving the brain puzzle. As noted above, there is no canonical method of data analysis and experimental design because the interpretation of functional brain imaging data requires assumptions that may be wrong. Optimal use of functional brain imaging will therefore rely on careful data mining of the raw data. This can be achieved by exploratory and Bayesian methods of data analysis, which are emphasized in this book. Also, there is no generic method of experimental design, and paradigms have to be designed tailor-made for the question to be answered. Furthermore, the temporal structure in brain activity provides indispensable clues about functional organization. At present, these can be assessed only by combining blood-based imaging with EEG/MEG methods. Approaches for combining these methods are described in the second part of this book. Finally, due to the massive underconstrainment of the system under consideration, mechanistic interpretations of macroscopic functional correlates depend on additional constraints that are not provided with the imaging data. These constraints may be results from microscopic functional studies, from neuroanatomy, or they may be computational assumptions infered from neural network studies. Linking such constraints requires network modeling of the imaging data, which is subject of the third part of this book. All told, this book is a valuable source of information not only to theorists in the field of neuroimaging, but to all experimenters striving for the best possible use of the brain imaging techniques to creatively address specific questions about the function of the human brain.

Manfred Spitzer
Ulm 2002

References

D. J. Stein and J. Ludik (eds). Neural networks and psychopathology. Cambridge University Press, 1998.

M. Spitzer. The mind within the net. MIT Press, Cambridge MA, 1999.

S. A. Huettel, P. B. Mack, and G. McCarthy. Perceiving patterns in random series: dynamic processing of sequence in prefrontal cortex. Nature Neuroscience 5, 485-490, 2002.

This page intentionally left blank

Preface

This book is a result of a workshop about theoretical methods in neuroimaging that took place in December 2000 in Breckenridge, Colorado. The workshop was part of the Neural Information Processing Systems (NIPS) conference, an annual interdisciplinary event that brings together cognitive scientists, computer scientists, engineers, neuroscientists, physicists, statisticians, and mathematicians interested in all aspects of neural processing and computation.

The purpose of this book is to provide a survey of theoretical and computational approaches in neuroimaging, communicated in the thin air of high altitude, to the broader community of scientists interested in neuroimaging.

We thank all participants of the workshop for creating the momentum for this book, and all the authors for their contributions. We wish to thank the NIPS Foundation, which not only provided the forum for the workshop but also made the publication of this book possible. We thank Axel Baune, University of Ulm, and Qingbo Wang, University of Southern California, for solving baffling technical problems as they arose. Editing of this volume was made possible by a grant from the state of Baden-Württemberg and by support from Wilhelm-Schweizer-Zinnfiguren GmbH in Diessen.

Chapter 7 is reprinted with permission of The MIT Press from *Neural Computation* 14, no. 8 (Aug. 2002): 1827 $\simeq$ 1858.

This page intentionally left blank

1 Theories, Data Analysis, and Simulation Models in Neuroimaging—An Overview

Friedrich T. Sommer
Department of Neural Information Processing
University of Ulm, D-89069 Ulm, Germany
fritz@neuro.informatik.uni-ulm.de

Judith A. Hirsch
Department of Neurobiology
University of Southern California, Los Angeles, CA 90027, USA
jhirsch@usc.edu

Andrzej Wichert
Department of Psychiatry
University of Ulm, D-89075 Ulm, Germany
wichert@neuro.informatik.uni-ulm.de

Functional neuroimaging techniques provide novel and exciting means for the investigation of working brains. Successful implementation of these tools requires an understanding of how to incorporate and adapt existing empirical results and theoretical frameworks for the design and analysis of imaging studies. Development of new techniques for analyzing data from neuroimaging is also important. No single technique can optimize the amount of information that can be extracted from neuroimaging studies; a broad spectrum of theoretical approaches is required. This chapter gives a brief overview of theoretical methods that are central to the field of experimental neuroimaging. The topics discussed include inferential, exploratory and causal methods of data analysis, theories of cerebral function and both biophysical and computational models of neural nets. As well this section helps to guide the reader by refering to later chapters.

1.1 Functional Neuroimaging: Answers without Questions?

The total effort devoted to functional neuroimaging is so great that it has led to the formation a new scientific field, Human Brain Mapping or Brain Imaging, with large numbers of associated conferences and journals. This intense interest is rooted in the distinctive power of techniques such as functional magnetic resonance imaging (fMRI) and positron emmission tomography (PET), which provide noninvasive means of viewing global patterns of neuronal processing in the human brain with spatial resolution at the millimeter scale. After its recent introduction by Ogawa et al. (1990a,b), fMRI in particular, has had an enormous impact. In the fields of cognitive neuroscience and systems neuroscience, fMRI became "the" registration technique of choice for examining macroscopic activation correlates in the working brain (Cabeza and Kingstone, 2000).

While there is tremendous interest in functional brain imaging, there is lively debate about its ultimate value and use. A recent article states succinctly: "It is unclear that we will come to a better understanding of mental processes simply by observing which neural loci are active while subjects perform a task," (Kosslyn, 1999). The author holds the view that the context given by prior studies and theories of cerebral function is required to pose questions that can be investigated by imaging studies. Kosslyn (1999) discerns two classes of questions that can be easily addressed by neuroimaging: first, how is information processing implemented in the brain and second, what are the time courses of activation of particular structures and processes. The work and theories that Kosslyn draws on have come from electrophysiological experiments, studies of the effect of brain damage on human behavior and cognition, and neuroanatomical descriptions of connections between different regions. Thus, there is a demand to translate prior empirical results and associated theories to the new medium of brain imaging. That is, to ask how can these be represented in mathematically defined ways. In the following, we will characterize approaches to data analysis and modeling in the context of functional brain imaging.

1.2 Techniques of Functional Neuroimaging: Strengths and Limitations

First, we will briefly describe the substance of different neuroimaging methods. Specifically, it is important to recall that all current techniques measure local neuronal activity by indirect means. PET and fMRI measure local properties of the cerebral blood flow: the fMRI signal is based on blood oxygen level dependence (BOLD), and the PET signal on regional cerebral blood flow (RCBF); for an introductory review, see (Horwitz et al., 2000). The mechanisms that link metabolic measures to neural activation are not yet well understood. The general impression is that the BOLD signal reflects the magnitude of synaptic events more closely than that of firing rates (Jueptner and Weiller, 1995; Magistretti and Pelerin, 1999).

Recently, this view has received further experimental support by a study that combined fMRI with electrophysiological recordings in animal models (Logothetis et al., 2001). Despite the indirect linkage between the BOLD signal and spike rate, fMRI has sufficient spatial resolution to resolve elements of functional architecture that were originally defined by the spatial distribution of stimulus evoked single-cell activity, such as orientation columns in the visual cortex (Kim et al., 1999). At present fMRI achieves the best spatial resolution that is possible for whole brain imaging but it is still 4–5 orders of magnitude away from discriminating single cells. New experimental methods are being developed, however, for characterizing the spatial distribution of neuronal populations beyond the technical spatial resolution of fMRI, see the adaptation paradigm proposed in the chapter by Tolias and collegues.

Imaging methods that depend on hemodynamic coupling are not only severely limited in spatial but also in temporal resolution. Technically, fMRI can be sampled at intervals less than 100ms, however, the hemodynamic response is unlikely to convey changes in neuronal response on such brief time scales. For example, the BOLD response begins with a weak dip in the blood oxygen level (depletion dip) that lags the he neuronal response by 1sec. The most pronounced BOLD response, the overshoot in oxygen level doe not peak until about 5–7s after the neuronal event (Frahm et al., 1994). Therefore, fMRI and PET provide only a version of the neuronal response that is low-pass filtered by 6 orders of magnitude.

Other measures of brain function permit superb temporal resolution. For example, EEG (electroencephalogram) electrodes and MEG (magnetoencephalogram) sensors record the electric or magnetic field arising from neuronal activity and achieve temporal resolution on the millisecond scale. In the past, these fast techniques provided very poor spatial resolution since they rely on integrating signals from large areas, 1–2cm around a detector. To some extent, spatial resolution can be increased by using a larger number of sensors. In the past few years, the number of detectors that can be fitted on a 2D surface near the skull has increased steadily; the present limit is about 256 detectors. From the field distribution recorded by many detectors on the surface of the skull, it is under certain circumstances possible to perform "source reconstruction," that is, to localize the spots of high density of dendritic currents in the brain underlying the measured field distribution. Thus, current encephalogram methods can also be thought of as a form of neuroimaging. Recent approaches have made progress in improving the spatial resolution of EEG/MEG and in resolving principal problems with source reconstruction. To date, EEG/MEG recordings can achieve accuracy on a scale of about 1cm, but not the millimeter scale of fMRI. Subsequent chapters will describe new analytical methods to improve sensitivity and spatial resolution (see section 1.5) and approaches that combine the advantages of EEG/MEG recordings with those of fMRI (see section 1.7.2).

1.3 Theories of Brain Function

A theory of brain function is a teleological interpretation of experimentally defined brain states. The traditional experimental bases of functional brain theories fall into two broad categories: a) Lesion studies that assess how cerebral injuries or other manipulations effect function; b) Recordings with electrodes or microelectrodes that measure neuronal activity in response to peripheral stimulation or during the performance of tasks. Lesion studies, for instance, provided the initial basis for hypotheses about the localization of brain function – a particular class of functional theories that specify which brain regions are involved in producing particular functions. Lesions in the occipital cortex, for example, produce blindness (Munk, 1880). Recordings of neuronal activity with microelectrodes placed in the brain provided the basis for functional theories at the neuronal level. For instance, neurons in the occipital cortex were found to code particular low-level properties of the external visual world such as stimulus orientation (Hubel and Wiesel, 1962). A hypothetical function of the primary visual cortex is therefore the decomposition of visual scenes into stereotyped features.

Two general principles of cerebral function derive largely from studies of brain lesions and recordings from single or small groups of neurons: One is the concept of functional specialization of brain regions (Zeki, 1990). The second is the hypothesis of functional integration (Gerstein and Perkel, 1969; Gerstein et al., 1989), which states that cerebral functions are carried out by networks of interacting regions and that different functions correspond to different networks.

1.4 Data Analysis of fMRI and PET

Functional neuroimaging complements traditional avenues of brain research by opening a macroscopic window on processing in the working brain. Changes of activity associated with various stimulus conditions and behaviors are referred to as functional correlates in brain activity. The extraction of functional correlates from spatio-temporal fMRI or PET data sets requires the application of sophisticated data analysis. One main difficulty in resolving functional correlates is to separate these from various types of distortion present in the measured signal, e.g. low pass filtering, physiological and scanner noise. Another problem in interpretation stems from the possibility that functional correlates of brain activity may relate to given behavioral paradigms in complicated ways. This latter difficulty would remain even if all issues of signal distortion and filtering were solved. Delineating functional correlates in the spatio-temporal structure of the data cannot be done without making assumptions about general working principles in the brain. Currently, there are two main types of assumption underlying the interpretation of functional neuroimages, as represented by the subtraction paradigm and the covariance paradigm (Horwitz et al., 2000).

1.4.1 Data Analysis Paradigms

The subtraction paradigm assumes that different brain regions are engaged in different brain functions (Horwitz and Sporns, 1994); that is, it relies on the existence of functional specialization. The subtraction paradigm has become the standard in most fMRI and PET studies. These studies commonly employ an experimental protocol known as block design, which involves switching between two steady states, or blocks, one a rest interval and the other a functional condition. This simple alternation between control state and behavioral task constrains the temporal structure of functional components of the signal. Thus, the data can be subjected to a regression analysis to reveal functional activation at any given location, see section 1.4.3. Many issues, however, cannot be addressed directly with experiments that use block design protocols, including studies about continuous voluntary movements, self-paced tasks and various forms of cognitive tasks.

The second paradigm is called the covariance paradigm (Horwitz and Sporns, 1994). It is motivated by the hypothesis of functional integration. Covariance paradigms assess the temporal covariance between different brain regions during a particular task. Significant covariance between regions associated with a particular brain function is termed functional connectivity. Originally, functional connectivity was determined by seed methods, by establishing signal covariance in different brain regions with respect to a chosen seed region. Currently, exploratory analysis techniques allow assessment of functional connectivity without reliance on seed regions (exploratory data analysis is explained in section 1.4.4 of the chapter).

In order to resolve all the functional components of a given cerebral process available from an fMRI data set, it is often necessary to take advantage of the complementary views that the subtraction and the covariance paradigms provide. For example, areas that are activated during a particular task, but not exclusively activated, would be missed in studies that rely on the subtraction paradigm alone. On the other hand, if only one small region, rather than multiple sites, is activated during a given task, its functional role would be undetected by the covariance approach. Therefore, analyzing the data with both approaches is often necessary. So far, we have given only broad definitions of the two main types of data analysis. In the following four paragraphs we will characterize new strategies of signal preprocessing and analysis and explain how these approaches relate to the subtraction and covariance paradigms.

1.4.2 Data Preprosessing

The detection of functional correlates from fMRI data sets can be improved by using the subtraction or covariance paradigms for prepossessing. For example, the covariance paradigm (combined with the anatomical finding that local networks are interconnected), led to the suggestion that functional activation extends over more than a single unit of measurement, or voxel (typically 1mm×1mm×3mm). This realization, in turn, gave rise to various means of reducing noise in measurements

of spatially extensive functional correlates spatial averaging such as smoothing by convolution with a Gauss-kernel. Further approaches to averaging came to include selection of voxel sets by similarity in signal time-course rather than by spatial proximity. Partitions for such selective averaging approaches can be found by explorative data analysis methods, see section 1.4.4 and the chapter of McKeown.

1.4.3 Inferential Data Analysis

Inferential analysis tests hypotheses about functional correlates in data sets from neuroimaging. A general approach to inferential analysis is to use spatially extended processes, statistical parametric maps (SPM). The most established sort of SPM is the general linear model, including familiar methods like ANCOVA (regression analysis), correlation coefficients and t-test as special cases (Friston et al., 1995). Inferential analysis involves the use of hypotheses drawn independently from the data set under study. Some experimental paradigms make a straightforward suggestion for the independent hypothesis to use. For block designs with block durations several times longer than the time constant of the hemodynamic response (6s), the hypothetical time course of the functional correlate is given by a box-car function, high levels during task periods and low levels during rest periods. The hypothesis can be tested in the SPM framework for each voxel. The resulting map of t-values is a picture of the spatial distribution of functional activity induced by the task. Other experimental paradigms, however, might suggest several competing hypotheses, or none at all. Previously, Burock and Dale (2000) developed a voxel-based method to estimate activation functions from data sets in cases for which no hypothesis is available. They also proposed a statistical framework capable of testing the activation functions estimated from those data sets. For testing competing hypotheses a novel Bayesian approach is presented in the chapter by Hansen and colleagues.

1.4.4 Exploratory Data Analysis

Strategies of multivariate data analysis that rely on the covariance paradigm represent other types of approaches that are free of preassumptions about activation functions. For example, methods of unsupervised learning, like cluster analysis or principal/independent component analysis, are able to reveal voxel sets with covarying time courses. Such algorithms, combined with only few preassumptions, have the ability to detect regularities in data from neuroimaging. Many studies have demonstrated that functional activity can be detected without reference to the experimental protocol at all. Exploratory data analysis has the capacity to reveal other components in the data as well, including scanner and motion artifacts. Exploratory data analysis is a main focus of this book; the methods and their application for different imaging techniques are described by a number of chapters: fMRI: Samorjai and collegues, McKeown; event-related fMRI: Wichert and collegues; MEG/EEG: Tang and Barak, Vigario and collegues; autoradiography: Nair and Gonzalez-Lima.

1.4.5 Causal Data Analysis

The types of analysis we described above were developed to reveal statistical regularities in the data that can be associated with brain function. The next step is to explore the processes that produce functional correlates in the brain, for instance, how different coactivated or sequentially activated areas influence one another. These issues cannot be resolved by determining functional connectivity alone. If, during task execution, activation of one region is associated with that of the next, the two are described as functionally connected; the causes or nature of the association, however, is unspecified. To explore potential mechanisms of interaction among regions, the data must be analyzed anew. Instead of statistical inference, a different type of analysis, causal inference, becomes important. Standard inference assumes that parameters that describe a given distribution can be inferred from samples taken from that distribution. These parameters can be employed to infer associations among variables, like the BOLD signal or behavioral features, with methods such as regression analysis. Causal analysis (originally developed for effect analysis in economics) goes one step further, by providing the means to make inferences about the processes involved in generating the data. It cannot be applied in all cases, however, since it requires supplementary information about the mutual interactions among variables like knowledge of anatomical connections. For a comprehensive introduction to the general concept of causal inference, see (Pearl, 2002). McIntosh and Gonzalez-Lima (1991) were the first to apply causal analysis to functional neuroimaging. In particular, they used structural equation modeling (SEM), a linear version of causal analysis. Their analyses provided a way to quantify the influence that a given cortico-cortical pathway (already known to exist from anatomical studies) has on its target area by generating path coefficients[1]. High coefficients indicated strong "effective connectivity" and negative coefficients indicated inhibitory effects.

1.5 Data Analysis of EEG and MEG

Since electro- and magneto-encephalogram registration is a longer established technique than fMRI or PET, the various problems of associated data analysis have received much previous attention. Here we limit the discussion of this longer used technique to two recent methods of exploratory data analysis and the combination with data from blood-flow based neuroimaging techniques.

The main goal of EEG/MEG data analysis is source reconstruction, as already has been described in section 1.2. The result of source reconstruction is a configuration of sources in the brain (time-dependent electric or magnetic dipoles or multipoles)

1. For technical reasons a SEM analysis has to be restricted to networks including only a handful of areas of interest.

that reproduce the field measured on the cranial surface. The reconstructed sources correspond to local cerebral regions with high dendritic current densities. In general, however, source reconstruction has no unique solution and so falls into the category of ill-posed problems[2].

One approach to reduce ambiguity in source reconstruction is to decompose the data set in a sensible way and to try to explain the components separately. Some approaches decompose the data by means of the same exploratory data analysis techniques described in section 1.4.4. Recent approaches using independent component analysis of single trial EEG and MEG will be described in the chapters of Tang and Pearlmutter, and Vigario and collegues. Even after successfull source reconstruction the spatial resolution of EEG/MEG is 1cm or coarser, far worse than the spatial resolution of fMRI—it depends on the number and localization of sources.

Another approach to reduce ambiguity in source reconstruction is to combine EEG/MEG with other imaging methods. Since advantages and disadvantages of EEG/MEG on the one hand and blood-flow based methods on the other are complementary, the combination of MEG/EEG with blood-flow based imaging methods is particularly appealing. The hope is to achieve high temporal and spatial resolution at the same time. The problems with combining different neuroimaging methods and ways to overcome them will be addressed in two of the following chapters, see section 1.7.2 below.

1.6 Neural Network Models

Applying causal data analysis in neuroimaging bears a close relationship to approaches of brain modeling in the field of computational neuroscience (Sejnowski et al., 1988); for a recent review see the supplement to Nature Neuroscience (volume 3, 2000). A central class of models are neural networks, which, at various levels of detail or abstraction, describe the interactions among groups of neurons. Two different types of neural networks are of particular value in the context of neuroimaging, biophysical models and computational models.

1.6.1 Biophysical Models

Biophysical models are descriptions of biological domains in the usual sense of physics. In physics, a model provides a sketch; a simplified view of a domain, with the degree and quality of the simplification determined by the modelers. They

2. Mathematical problems are ill-posed if they do not satisfy each of three criteria: a solution exists, is unique, and depends continuously on the initial data. To solve ill-posed problems, well-posedness must be restored by restricting the class of admissible solutions (Hadamard, 1923).

select experimental phenomena, mechanisms, and interactions that are regarded as essential to describe and formulate the descriptions in mathematical language. The resulting mathematical model can be treated analytically, numerically, or implemented in computer simulations. The biophysical model can be used to predict how experimental parameters influence other ones. To assess the validity of the model, the predictions it makes can be tested experimentally. Faithful or valid models usually result from a recursive trial and modification process. A biophysical model should not only be predictive, it should be explanatory too. Explanation largely results from the simplification process, which helps to identify the most important mechanisms involved in biological processes, such as how local interactions effect macroscopic behavior. Fidelity and simplification/reductionism are equally essential to biophysical modeling. A model that does not faithfully represent the system it seeks to describe becomes invalid; as does one that fails to provide reductionistic explanations for the actions of that system. Of course, model simplification is dictated by pragmatic reasons as well; mathematical models have to be tractable and should not include parameters that cannot be measured empirically. Naturally, these pragmatic limitations change as mathematical tools and technology develop. For example, recent advances in computing power and new experimental techniques have expanded the scope of biophysical models. A classic example of a biophysical neural network model is one that draws on the properties of single neurons and their synaptic interactions to explain the behavior of the network as a whole (Pinsky and Rinzel, 1994; Traub et al., 1996). Other biophysical models describe how extracellular field potentials, measured by EEG/MEG, are produced by postsynaptic potentials of large groups of neurons (Nunez, 1990).

1.6.2 Computational Models

The cybernetics movement introduced the computational paradigm of brain function, that is, the idea that the role of the brain is computation. This movement initiated the search for algorithmic formalizations of cerebral functions (Craik, 1943; Rosenblueth et al., 1943; Wiener, 1948). Just as biophysical models describe neuronal domains, algorithms provide mathematical descriptions of cerebral functions. Cybernetics eventually gave rise to the field of Artificial Intelligence (AI), whose goal is to describe cognitive brain functions with mathematical algorithms. The rationale of this synthetic approach was that biological functions might be easier to mimic than to analyze in situ. Thus, algorithms implemented in a technical system (a computer) yield predictions that can be compared with the performance of biological systems. The hope was that if artificial and natural behaviors were similar, algorithmic elements would help to define the biological working principles. While the AI approach is interesting, its use in specifying the mechanisms of biological behavior is limited by the fact that different algorithms can produce identical outcomes. Computational brain models share with AI the goal of reproducing brain functions in an artificial system. But in addition, the structures and processes in the models are constrained by biophysics. Thus, neural networks used as computa-

tional models link interfaces to two different aspects of the brain, its biophysics and the algorithms it implements. Algorithms performed by neural networks that are important for neuroimaging include, for instance, associative memory, the storage and recall of activity patterns.

1.7 Neural Network Models for PET and fMRI

1.7.1 Links to Multielectrode Recordings

Biophysical models developed to describe functional activity as measured by PET and fMRI have been designed (Arbib et al., 1995; Tagamets and Horwitz, 1999; Horwitz and Tagamets, 1999) to bridge the gap between results from microelectrode recordings of single cells and imaging of the whole brain. At present, these models are selective in scope because the path between single cell recording and functional imaging is long and only loosely charted. These models have been constructed to include information about various connectivity schemes between different local neuronal populations and between the areas of interest. Initially the models were used for simulations that explored the role of inhibitory and excitatory neuronal populations and the biophysical relationship between blood flow and neural activity. An important future role for the simulations will be to plan novel neuroimaging experiments. (see chapters Arbib and collegues, Tagamets and Horwitz.).

1.7.2 Links between Different Neuroimaging Techniques

Blood flow based and electrophysiological measurements have complementary strengths and weaknesses in space and time; thus experiments that combine both approaches are potentially powerful. Before these approaches can be brought together, however, two main obstacles must be overcome. First, technical solutions must be found for solving interference problems caused by the conjoint application of both techniques (see chapter Kruggel and Hermann). Second, a means to combine the signals measured with the two techniques must be developed. The approach to this second problem is not clear cut. Indeed, it is not even known whether both signals arise from the same or different groups of cells (Nunez and Silberstein, 2000). As well, there is considerable debate about which properties of the EEG signals (e.g., ERP peak height, different bands of spectral power) correlate best with the fMRI BOLD signal (see chapter Makeig and collegues).

1.7.3 Refined Causal Data Analysis

The strength of causal data analysis of fMRI/PET has recently been improved by the inclusion of biophysical models. Early methods of causal data analysis reflected only rudimentary descriptions of the biophysical substrate for interactions between regions. More recent methods of causal data analysis use fuller descriptions of

biophysical properties, including populations of inhibitory as well as excitatory neurons and the relationship between blood flow and neuronal activity (Taylor et al., 2000). Moreover, principles from computation modeling, like pattern storage or hierarchical processing have also been implemented. These revised analyses determine the influences brain regions exert on each other more realistically than previous causal analyses had done.

1.7.4 Neuroimaging Studies on Learning

Learning is one of the most exciting subjects of studies that can be addressed by causal data analysis. One common computational model predicts that learning initiates changes in synaptic connectivity. Hence, learning might induce changes in effective connectivity as well. In fact, associative learning of visual objects and their location was shown to influence effective connectivity between regions specialized for spatial and object recognition (Büchel et al., 1999). Developmental influences on such learning effects is the subject of an autoradiographic study of neural tissue in the chapter of Nair and Gonzalez-Lima.

1.8 Final Remarks

This overview gave brief accounts of different approaches to neuroimaging, including methods of analysis, modeling and the development of experimental techniques. In addition, we tried to convey a sense of how the different approaches relate to one another. For instance, there are close relationships between causal data analysis and biophysical modeling, and between functional theories and computational models. Last, by discussing biophysical and computational models separately, we hoped to clarify the different roles that neural network models play in understanding and integrating different types of information about the working brain.

References

D. J. Amit. *Modelling Brain Function*. Cambridge University Press, Cambridge, 1989.

M. A. Arbib, A. Bischoff, A. Fagg, and S. A. Crafton. Analyzing large-scale properties of neural networks. *Human Brain Mapping*, 2:225–233, 1995.

C. Büchel, J.T. Coull, and K.J. Friston. The predictive value of changes in effective connectivity for human learning. *Science*, 283:1538 – 1541, 1999.

M. A. Burock and A. M. Dale. Estimation and detection of event-related fmri signals with temporally correlated noise: A statistically efficient and unbiased approach. *Human Brain Mapping*, 11(4):249–260, 2000.

R. Cabeza and A. Kingstone, editors. *Handbook of functional neuroimaging of cognition*. MIT Press, Cambridge, MA, 2000.

K. J. W Craik. *The nature of explanation*. Cambridge University Press, Cambridge, UK, 1943.

J. Frahm, K. D. Merboldt, W. Hänicke, A. Kleinschmidt, and H. Boecker. Brain or veinoxygenation or flow? on signal physiology in functional MRI of human brain activation. *NMR in Biomedicine*, 7:45–53, 1994.

K. J. Friston, A. P. Holmes, K. J. Worsley, J. B. Poline, C. D. Frith, and R. S. Frackowiak R.S. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2:189–210, 1995.

G. L. Gerstein, P. Bedenbaugh, and A.M.H.J. Aertsen. Neuronal assemblies. *IEEE Transactions of Biomedical Engineering*, 36:4–14, 1989.

G. L. Gerstein and D. H. Perkel. Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science, Wash. DC*, 164:828–830, 1969.

J. Hadamard. *Lectures on the Cauchy Problem in Linear Partial Differential Equations*. Yale University Press, Yale, CT, 1923.

D. O. Hebb. *The Organization of Behaviour*. Wiley, New York, 1949.

B. Horwitz, K. Friston, and J.G. Taylor. Neuronal modeling and functional brain imaging: an overview. *Neural Networks*, 13(8-9):829–846, 2000.

B. Horwitz and O. Sporns. Neural modeling and functional neuroimaging. *Human Brain Mapping*, 1:269 – 283, 1994.

B. Horwitz and M.-A. Tagamets. Predicting human functional maps with neural net modeling. *Human Brain Mapping*, 8(2):137–142, 1999.

D.H. Hubel and T.N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 160:106–154, 1962.

M. Jueptner and C. Weiller. Review: Does measurement of regional cerebral blood flow reflect synaptic activity?-implications for PET and fMRI. *NeuroImage*, 2 (2a):148–156, 1995.

D.-S. Kim, T.Q. Duong, and S.-G. Kim. High-resolution mapping of iso-orientation columns by fMRI. *Nature Neuroscience*, 5(2):164–169, 1999.

S. M. Kosslyn. If neuroimaging is the answer, what is the question? *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 354:1283–1294, 1999.

N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412:150–157, 2001.

P. J. Magistretti and L. Pelerin. Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. *Philosophical Transactions of the Royal Society of London B*, 354:1155–1163, 1999.

A. R. McIntosh and F. Gonzalez-Lima. Structural modelling of functional neural pathways mapped with 2-deoxyglucose: Effect of acoustic startle habituation on the auditory system. *Brain Res.*, 547:295–302, 1991.

H. Munk. über die sehsparen der grosshirnrinde. *Monatsberichte K. Akademie d. Wissenschaften*, pages 1–23, 1880.

P. L. Nunez. Physical principles and neurophysiological mechanisms underlying event-related potentials. In J. W. Rohrbough and R. Parasuraman, editors, *Event-related brain potentials*, pages 19–36. Oxford University Press, New York, 1990.

P. L. Nunez and R. B. Silberstein. On the relationship of synaptic activity to macroscopic measurements: Does co-registration of EEG with fMRI make sense? *Brain Topography*, 13:79–96, 2000.

S. Ogawa, T. M. Lee, A. S. Nayak, and P. Glynn. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Journal of Magnetic Resonance in Medicine*, 14:68–78, 1990a.

S. Ogawa, T.M. Lee, A.R. Kay, and D.W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences, U.S.A.*, 87:9868–9872, 1990b.

J. Pearl. Causal inference in the health sciences: a conceptual introduction. *Health Sciences and Outcomes Research Methodology*, to appear in: Secial issue on causal inference, 2002.

P. F. Pinsky and J. Rinzel. Intrinsic and network rhythmogenesis in a reduced Traub model for CA3 neurons. *Journal of Computational Neuroscience*, 1:39–60, 1994.

A. Rosenblueth, N. Wiener, and J. Bigelow. Behavior, purpose and teleology. *Philosophical Sciences*, 10:18–24, 1943.

T. J. Sejnowski, C. Koch, and P. S. Churchland. Computational neuroscience. *Science*, 241:1299–1306, 1988.

M. Tagamets and B. Horwitz. Functional brain imaging and modeling of brain disorders. *Prog Brain Res*, 121:185–200, 1999.

J.G. Taylor, B. Krause, N.J. Shah, B. Horwitz, and H.-W. Mueller-Gaertner. On the relation between brain images and brain neural networks. *Human Brain Mapping*, 9(3):165–182, 2000.

R. D. Traub, M. A. Whittington, I. M. Stanford, and J. G. R. Jefferys. A mechanism for generation of long-range synchronious oscillations in the cortex. *Nature*, 383: 621–624, 1996.

N. Wiener. *Cybernetics: or control and communication in the animal and the machine*. MIT Press, Cambridge, MA, 1948.

S. Zeki. Functional specialization in the visual cortex: the generation of separate constructs and their multistage integration. In G. M. Edelman, W. E. Gall, and W. M. Cowan, editors, *Signal and Sense*, pages 85–130. Wiley, New York, 1990.

This page intentionally left blank

I fMRI Data Analysis and Experimental Designs

This page intentionally left blank

2 Exploratory Analysis of fMRI Data by Fuzzy Clustering—Philosophy, Strategy, Tactics, Implementation

Ray L. Somorjai
Institute for Biodiagnostics
National Research Council Canada
435 Ellice Ave.
Winnipeg, MB R3B 1Y6, Canada
Ray.Somorjai@nrc.ca

Mark Jarmasz
Institute for Biodiagnostics
National Research Council Canada
435 Ellice Ave.
Winnipeg, MB R3B 1Y6, Canada

The purpose of Exploratory Data Analysis (EDA) is to investigate and discover salient and novel features of complex, high-dimensional data. We describe a particular realization of EDA, the three-stage strategy EROICA (Exploring Regions Of Interest with Cluster Analysis), specifically designed to analyze functional MR neuroimaging data. The first stage consists of an *Initial Partition* of the data into three groups: a group of "trend" time-courses (TCs), a group of "potentially interesting" TCs, and a group that contains the remaining, putative "noise" TCs. The initial grouping is achieved by first normalizing (scaling) the TCs, followed by *selection* procedures based on specific "trend" and "noise" tests. The second stage is the *Principal Partition*, where fuzzy clustering analysis (FCA) is applied to the group of "potentially interesting" TCs. The third stage, *Significance Testing*, "validates" the second-stage results by first removing those TCs from the original clusters that fail special statistical tests, and then by attempting to allocate to the clusters some of the initially excluded "trend" and "noise" TCs. We assessed the consequences of this three-stage strategy on the quality of the clustering results. We show that employing this strategy both improves results relative to clustering that did not use

the initial partitioning, and also speeds up execution significantly. We report detailed analyses on several phantom datasets and on a multi-slice, real fMRI dataset. Based on detailed studies carried out on sixteen fMRI datasets, the execution time of EROICA scales sublinearly both with T (scans) and N (number of TCs). We propose robustness (noise resistance, reproducibility) flexibility/versatility, and speed as the three major requirements that any practically viable EDA method ought to satisfy. We show that the EROICA process, and EvIdent®, its software implementation, fulfill these requirements.

2.1 Introduction

Understanding how the human brain works has been an age-old quest. The advent of functional neuroimaging promises fulfillment. Different imaging modalities probe differently the brain's responses to stimuli. They all have advantages and disadvantages. For instance, the exquisite temporal resolution attainable by EEG (electroencephalography), MEG (magnetoencephalography) and ERP (event related potentials) is offset by the poor spatial (neuroanatomical) resolution. In addition, the so-called "inverse problem" has to be solved to disentangle and locate the sources from which the responses are collected at the electrodes.

The other, most commonly used modalities do not depend directly on neural responses, but rely on exogenous or endogenous probes that reflect only indirectly the actual neuronal activity, (generally through its coupling to metabolic activity). Thus, PET measures changes in regional cerebral blood flow, by following the time-course of a bolus of injected $H_2^{15}O$. Because the half-life of ^{15}O is 123 s, 6–12 scans, each lasting approximately 1 minute, can be performed in the same scanning session. However, the spatial resolution of PET images is poor, 5–6 mm.

Functional MRI (fMRI) is based on the BOLD (blood oxygenation level-dependent) effect, which assesses the changes in blood oxygenation and blood volume resulting from changes in neural activity. Deoxygenated hemoglobin (dHb) acts as an endogenous paramagnetic contrast agent. Increased blood flow reduces the local concentration of dHb, resulting in increased MR signal on a $T2^*$-weighted image. Both the temporal and spatial resolution of fMRI are much better than for PET, the former as high as 100 ms (EPI), the latter approximately 2 mm. A limitation of fMRI is that the hemodynamic response to a change in brain state is delayed by 5–8 s. However, fMRI, because of its non-invasive nature, excellent spatial and good temporal resolution, has become the neuroimaging method of choice.

The simplest and earliest data analysis strategy of neuroimaging experiments (especially PET) used the so-called *subtraction paradigm* (Posner et al., 1988). This relied on the notion of *functional specialization*, i.e., that different functions (stimuli) activate different regions of the brain. For a typical experimental paradigm, this involves the comparison of, say, two different experimental conditions. If statistically significant signal differences ("activations") between these conditions

can be located in particular brain regions, then the brain areas "activated" are presumably related to the differences between the two conditions.

Another early strategy is motivated by a more distributed view of cortical function, with the premise that the response to an experimental task is mediated by a network of interacting brain regions, different tasks by different functional networks. This assumption of distributed functions led to the so-called *covariance paradigm* (Horwitz, 1994). It postulates that by studying how brain activity covaries between different brain areas, one can infer which areas correspond to important network nodes, as well as the functional connectivity of these nodes.

The covariance paradigm assumes that there is strong correlation amongst the TCs of an activated area but that different areas correlate weakly. However, this assumption doesn't allow for strong correlation between spatially distributed areas that follow the same task paradigm, even though this is a common occurrence.

Both the subtraction and covariance paradigms use minimal assumptions (e.g., that there be two different experimental conditions). Interestingly, the focus of analysis shifted to inferential, confirmatory statistical approaches, especially because the tasks of early fMRI experiments were designed to be periodic (block design). The standard statistical inferential methods of analysis, based e.g., on Pearson's product-moment correlation coefficient ρ, or on the ubiquitous Student's t-test, are appropriate if a realistic and faithful model function of the expected response is known. More powerful statistical approaches, such as the generalized linear model (McCullagh et al., 1989), adapted by Friston and co-workers (Friston, 1995; Friston, 1996; Friston et al. 1995) for analyzing PET and more recently fMRI data, are satisfactory for more complicated paradigms, but still require modeling. However, selecting the most appropriate model, without additional information and further assumptions, is difficult or even impossible when the temporal response to a presented stimulus is complex or poorly characterized. A priori modeling is in principle impossible when the stimulus to be identified is spontaneous, endogenous, and/or non-generic (e.g., the task is to follow the behavior of a patient with Tourette's syndrome (Gates et al., 2001), or the onset and course of epileptic seizures, or the consequences of drug therapy, etc.).

Hence, the need for exploring neuroimaging (particularly fMRI) data by "model-free," data-driven methods is becoming more compelling as neuroscientists design increasingly sophisticated and probing cognitive/linguistic experiments, and as more and more potential clinical applications emerge. Real-life, i.e., complex, large, and feature-rich data, which do not have "nice" statistical properties such as homogeneity, stationarity, stochasticity, etc., cannot be reliably analyzed with the more commonly used, inferential, confirmatory data analysis (CDA) methods. Eminent statisticians (see e.g., Tukey, 1962) recognized this, and the recommended strategy and approach was formally enunciated (Tukey, 1977) as *exploratory data analysis* (EDA), a natural complement for CDA.

Attempts to analyze fMRI data by model-free EDA methods, such as Principal Component Analysis (PCA) (Jackson, 1991; Friston, 1996) or its nonorthogonal variant, factor analysis (FA) (Backfrieder et al. 1996), although occasionally suc-

cessful, failed just as often. The main reason for failure is that both PCA and FA segregate the data by partitioning its total variance into uncorrelated components. For PCA, the partitioning is along the mutually orthogonal PC axes. FA involves additional, oblique rotations of the PC axes. This removes the orthogonality constraint, without necessarily improving variance partitioning. The latter, whether achieved by PCA or FA, cannot always separate the data unambiguously into, say, activated time-courses (TCs), noise TCs, and TCs containing artefacts. This is because the expected amplitude of the activations is small, and variance partitioning is non-specific: thus, activated TCs are likely to be contaminated with, or even swamped by noise, whether instrumental or physiological (respiratory, cardiac, etc.). Spatial Independent Component Analysis (ICA) (Bell et al., 1995), recently applied to the analysis of fMRI data (McKeown et al., 1998), self-organizing neural nets(Fischer and Hennig, 1999; Chuang et al., 2000), and especially temporal Cluster Analysis (CA), are alternative unsupervised pattern recognition approaches that do not suffer from some of the disadvantages of PCA or FA. ICA is conceptually similar to PCA/FA, with the important difference that variance partitioning is based on mutual information (i.e., on higher than second-order correlations). One of the consequences of this difference is that the variance partitioning achievable is not "greedy," i.e., ICA, unlike PCA, does not necessarily try to maximize the variance in the data along successive orthogonal directions. However, ICA imposes the constraint of (spatial and/or temporal) statistical independence, often satisfied only approximately. Additional possible limitations are ICA's *linearity* (McKeown and Sejnowski 1998), and in particular, that, just as PCA/FA, it attempts to characterize the data *globally*. This means that even if the dataset is statistically heterogeneous (i.e., the time-courses in different regions of the brain have different distributional properties, certainly the case for fMRI data), the ICA model tries to describe it using the same *global* features (Karhunen and Malaroiu, 1999), i.e., as if the data were spatially homogeneous. In principle, some nonlinear version of ICA, expressed in various forms (Parra et al., 1996), might solve this problem of distributional heterogeneity. However, excessive computational requirements for high-dimensional data, and the likelihood of non-unique solutions, cast serious doubt on the practical realization of such nonlinear variants (Lin et al., 2001; Yang et al., 1998; Taleb and Jutten, 1997). (These reservations also apply to recent proposals of using some variant of nonlinear PCA, e.g., in (Friston et al., 2000).)

A typical 3-dimensional (multislice) fMRI brain dataset for a *single* subject contains about 10^8–10^9 bytes. According to Huber's data size taxonomy (Huber, 1994), such sizes are crudely classifiable as between "large" (10^8) and "huge" (10^{10}). *Simultaneous* inter-subject studies would raise the dataset size to "huge". The relevance of this size-based classification is that real-life "large" datasets differ from smaller ones not only by size. In particular, nonstationarity and especially distributional heterogeneity, two persistent features characterizing massive datasets, will likely play havoc with conventional statistical methods. Expecting or requiring a meaningful description of the data, by some *global* transformation, such as linear or nonlinear PCA or ICA, is unrealistic.

Clustering, particularly fuzzy clustering (FC) (Bezdek, 1981), does not suffer from the constraints implicitly imposed on ICA. It naturally represents the data both locally (cluster centroids rarely involve the combination of *all* TCs) and in a nonlinear manner. Furthermore, its algorithmic implementation can be made very fast, a feature we feel important if the goal is the thorough exploration of "large" (Huber, 1994) and complex datasets. The need for adequate computational speed when analyzing such datasets is convincingly argued in (Huber, 1994; Huber, 1997; Huber, 1999), (Wegman, 1995; Wegman, 2000; Wegman, 2000).

Since its introduction in neuroimaging in 1995, FCA/EvIdent was successful in analyzing brain fMRI data, both by us (Baumgartner et al., 1999; Baumgartner et al., 2000, Baumgartner et al., 1997), (McIntyre et al., 1998), (Scarth et al., 1995; Scarth and Somorjai, 1996), (Somorjai and Jarmasz, 1999; Somorjai et al., 1997; Somorjai et al., 2001; Somorjai et al., 1999), and by colleagues and collaborators (Barth et al., 1999), (Baumgartner et al., 2001), (Carpenter and Just, 1999), (Kato et al., 1999), (Moser et al., 1999; Moser et al., 1997; Moser et al., 1996).

The usefulness of FCA is not confined to brain fMRI. It was applied successfully to *perfusion* MRI of the brain (Ye et al., 1997), breast (Scarth et al., 1997) or heart (Tian et al., 1998), and to infrared image analysis of the skin (Mansfield et al, 1997), tissue oxygenation (Sowa et al., 1997) and viability (Mansfield et al., 1998) etc.

A critical, in-depth comparison of the various data-driven methods used in fMRI data analysis has not yet been made, and is beyond the intended scope of this Chapter. We confine ourselves to some of the conceptual and computational aspects of FC analysis, in the light of general EDA considerations. It has become obvious that the strategy of analysis and its implementation both play essential and complementary roles in creating a practically useful EDA method. Crucial requirements for EDA methods are 1) robustness (noise resistance, reproducibility), 2) flexibility/versatility, 3) speed and 4) easy applicability. The last condition implies a user-friendly GUI.

Our main purpose is to describe and discuss a multi-stage strategy, based on the concept of "divide and conquer," which achieves the above goals. (Earlier, we have announced a two-stage version (Somorjai et al., 1999; Somorjai et al., 2001). A simplified, more basic variant of this, using only the autocorrelation-based "self-similarity" (Somorjai et al., 2001), no trend exclusion, and the slower, original Bezdek fuzzy clustering algorithm, without cluster merging (Jarmasz and Somorjai, 1998), was adopted and implemented in (Fadili et al., 2000). The current three-stage strategy is an extension and improvement of our earlier version. It explicitly emphasizes the importance of testing the generated hypotheses for statistical significance.

We show that by appropriately *preprocessing* the data (Stage I), we can substantially enhance the success of fuzzy clustering in partitioning multidimensional vectors (the T-dimensional TCs in the case of fMRI) into functionally and/or physiologically meaningful groups. Typically, only a small fraction of the total number of TCs is activated by the task; hence, once the limited subset containing these TCs has been identified, FCA needs only to be applied to this subset. When "po-

tentially interesting" clusters have been found, we have to ensure that the initially excluded TCs are properly assigned to the most similar cluster. This motivated our developing the three-stage strategy/process we call EROICA (**E**xploring **R**egions **O**f **I**nterest with **C**luster **A**nalysis) (Jarmasz and Somorjai, 2002). It is important to establish that the initial partitioning is not at the expense of the quality of the results but, in fact, quality will also improve. Any consequent speedup in processing is a welcome bonus, and enables the analyst to more thoroughly explore the data. The flow-chart in figure 2.1 depicts the basic components of EROICA's three stages, with their functional interrelationships. These were implemented in the software EvIdent®.

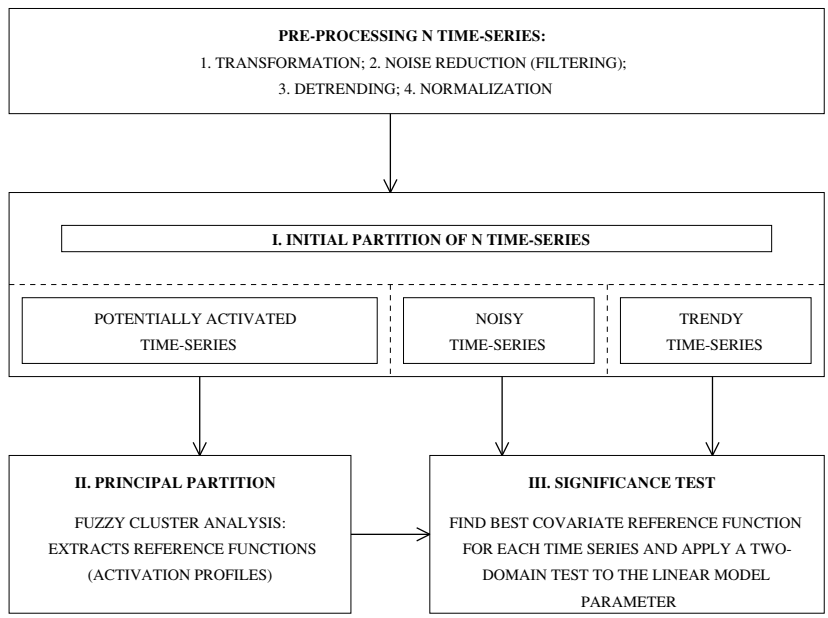


Figure 2.1 Flow-chart of EROICA, the 3-stage EDA strategy/process.

2.2 Methodology

Stage I: Initial Data Partition (Preprocessing)

If the inputs are the raw, unprocessed TCs, and the Euclidean metric is chosen as the distance measure, then the FC algorithm merely *segments* the brain (Scarth et al., 1995), i.e., TCs are assigned to clusters, based only on the magnitudes of their average intensities. Normally, this is not the goal of temporal FCA, which typically attempts to partition the TCs according to their temporal profiles (waveforms). To achieve this, the first step of preprocessing is what we call *normalization* in the initial publication (Scarth et al., 1995). Its principal role is to remove from the data

the influence of different average intensity levels, and highlight temporal waveform shapes. The various possible normalization methods are discussed in the Appendix.

Following normalization, a specific TC *screening* method is chosen to carry out the initial data partition. The goal is to eliminate from subsequent FC analysis the confounding effects of noise and/or trend TCs and analyze only the "potentially interesting" ones. Both "trend" and "noise" tests (see fig. 2.1) were incorporated into EROICA (Jarmasz and Somorjai, 2002), to give it the robustness (Requirement 1) and speed (Requirement 3) that any practically useful EDA method should have. The following are concise descriptions of the preprocessing methods we found most useful. The subsequently defined test procedures make extensive use of the Pearson product-moment correlation coefficient ρ ($-1 \leq \rho \leq 1$), defined as

$$\rho(\mathbf{x}_i, \mathbf{v}_j) = \frac{\mathbf{x}_i^t \mathbf{v}_j}{\sqrt{\mathbf{x}_i^t \mathbf{x}_i \mathbf{v}_j^t \mathbf{v}_j}} \tag{2.1}$$

where $\mathbf{x}_i$ and $\mathbf{v}_j$ are zero-mean T-scans vectors, and the superscript "t" denotes the *transpose*. While the test A) is used to select voxels with distorted signal, the tests B) and C) can identify "potentially interesting" TCs.

A) *Trend Exclusion (TE)*: Prior to clustering, we detect TCs with significant trends, and *temporarily* place these in a "trend" cluster (see figure 2.1). (Trends may be due to motion artefacts and/or instrumental drift (Smith et al., 1999). In EPI data often as many as $50 - -70\%$ of the voxels have statistically significant trends.)
We have implemented trend detection/exclusion as a two-stage process:

- We compute the Pearson correlation coefficient ρ between each TC and a straight line with unit slope; if $|\rho| \geq \rho_0$, the TC is used in creating a "trend" centroid with ρ as weight. The correlation threshold is obtained from $\rho_0 = \sqrt{2\frac{SP_0(m)}{T-1}}$, where the Spectral Peak threshold $SP_0(m) = -ln[1-(1-p)^{1/m}]$, p is a user-defined false-positive rate, and m is the number of spectral peaks in the power spectrum density of a straight line with unit slope that are above the lowest significance level of $SP_1(m) = -ln(p)$ (see (Jarmasz and Somorjai, 2002) for more details). p determines the level of statistical significance (confidence level) - the smaller p the less likely that a match with the straight line is due to chance. All TCs with $|\rho| \leq \rho_0$ are averaged, with weights set equal to the ρ values, to create a "trend" centroid. We will give a more comprehensive explanation of the origin of $SP_0(m)$ and ρ_0 in the Significance Test section.
- All TCs are now correlated with the "trend" centroid. TCs with $|\rho| \geq \rho_0$ are placed in a "Trend" cluster and are excluded from the Principal Partition.

Once the Principal Partition is completed, an attempt is made to reassign trendy TCs to the clusters found by the FCA. The criterion for reassignment will be explained in the Significance Test section. In our experience, most of trendy TCs tend to remain in the "Trend" cluster. Clearly, trend exclusion is a *data-driven*

process, and the "trend" centroid C_{trend} is created by the data. Often C_{trend} has a highly nonlinear shape. We find that temporary trend exclusion is generally superior to trend removal (i.e., "detrending," usually accomplished by fitting TCs to a low-degree polynomial), since detrending may create spurious temporal shapes in TCs that in fact have nonlinear trends, a very common situation. The distinction between trend exclusion and trend removal (that actually alters the TC) is important. We do not permanently detrend TCs, and no TC is permanently rejected without an appropriate statistical test. The temporary exclusion of trend TCs from analysis not only speeds up subsequent processing; it makes the process of finding "interesting" cluster centroids more robust (Somorjai and Jarmasz, 1999).

B) *Autocorrelation (AC)* (Somorjai et al., 1999): This method is based on the observation that a potentially activated (i.e., structured) TC that is shifted by one time instance (lag) has a high correlation coefficient, designated by AR(1), with the original, unshifted TC (i.e., in equation (2.1) we set $\mathbf{v}_j = \mathbf{x}_{(i+1)}$). On the other hand, a time-shifted noise TC has consecutive amplitude values that are effectively independent of each other, and generates AR(1) $\cong 0$. AR stands for the *auto-regressive* modeling of a time-series. For Gaussian noise, the expected value of AR(1) is zero. The statistic AR(1) can therefore be used for detecting noisy TCs, i.e., all TCs that fail the hypothesis $AR(1) \leq AR_0(1)$ are potentially activated TCs that can be clustered independently of the noisy TCs. (Note that "trend" TCs also generate large AR(1) values, and must be excluded or corrected prior to selection.) For a sufficiently large T, the statistic $Q_1 = \frac{AR(1)^2 T(T+2)}{(T-1)}$ has a chi-square distribution with one degree of freedom (Ljung and Box, 1978). For a given false-positive p, the corresponding correlation threshold is given by $AR_0(1) = \sqrt{Q_1(p)(T-1)/T(T+2)}$, where $Q_1(p)$ is the inverse chi-square distribution. Only TCs with $AR(1) > AR_0(1)$ are selected for the second, Principal Partition stage. TCs with $AR(1) \leq AR_0(1)$ are placed in the "Rejects" cluster. Analogously to the trendy TCs, rejected TCs can be assigned to one of the clusters found by FCA if the membership threshold is met. The reassignment scheme is described in the Stage III (Significance Testing) section.

C) *Spectral Peak (SP)* (Jarmasz and Somorjai, 2002): This method is based on a frequency-domain solution to a common signal-analysis problem: how to detect the presence of a periodic (or nearly periodic) signal that is buried in noise. We describe a spectral peak order statistic we use to identify potentially activated TCs prior to clustering. Most activated TC can be modelled as a sum of one or two dominant periodic signals plus noise. For such signals, the peak in the power spectral density (periodogram) makes a disproportionately large contribution to the total signal power. Thus, the actual fraction of total signal power contained in a given spectral peak can be a useful measure in identifying TCs dominated by one or two spectral peaks. For noisy TCs, the total power is approximately evenly distributed over the entire spectrum, and this leads to a comparatively small spectral peak measure. We define a spectral peak statistic SP, as the power contained in

the spectral peak divided by the average power. (SP is a scaled version of Fisher's g statistic (Fisher, 1929).) To compute the SP statistic, the TCs are padded with zeros so that $N = 2^n > 2T$, where T is the number of scans (time instances), and the integer part of [(N-1)/2] is the number of frequency points that are searched to find the peak. (To allow proper statistical analysis, we ignore the Nyquist frequency N/2.) Padding with zeros increases frequency resolution and letting $N = 2^n$ permits using an efficient FFT algorithm to compute the power spectrum (this is an important step for minimizing execution time). For TCs drawn from a Gaussian process, the probability distribution for the statistic was derived by Fisher (Fisher, 1929) and its first term is given by $Pr\{SP > SP_0\} = M(1 - SP_0/M)^{M-1}$, where M = integer part of [(T-1)/2]. For a given false-positive rate (significance level) p, the SP threshold is given by $SP_0 = M[1 - (p/M)^{1/(M-1)}]$. All TCs with $SP > SP_0$ (i.e., those that fail the noise test) are deemed "potentially interesting" and are selected for the Principal Partition; the others are placed temporarily in the "Rejects" cluster (see fig. 2.1).

Stage II: Principal Data Partition—Fuzzy Cluster Analysis

The goal of Stage I, *Initial Partition*, is to find a sufficient number of potentially activated TCs such that they can form valid clusters at the second, *Principal Partition* stage. Fuzzy clustering, the second stage of EDA in EROICA, starts with selecting a *distance (similarity) measure* for assessing how similar two TCs are to each other. There is clearly nothing unique about the Euclidean metric as the similarity measure. The choice among an infinite number of possibilities, provided e.g., by the well-known Minkowski distance metric D_{ij}^{λ}:

$$D_{ij}^{\lambda} = \{\sum_{k=1}^{T} |x_{ik} - v_{jk}|^{\lambda}\}^{1/\lambda}, \quad -\infty \leq \lambda \leq \infty$$

must be dictated by extrinsic considerations. ($\lambda = 2$ produces the familiar Euclidean metric, $\lambda = -\infty$ selects the smallest, $\lambda = \infty$ the largest of the T terms in the sum.) A good discussion of other possible measures is found in (Goutte et al., 1999). In quite general terms, similarity measures of any kind (Gower, 1971) could be used, with "similarity" and "distance" related by an appropriate (and possibly quite arbitrary) monotonically decreasing transformation. (Note however, that for each new measure, the validity/ convergence properties of the alternating-stage iterative algorithm used to carry out the optimization in FC must be verified (Bezdek et al., 1999).)

(Golay et al., 1998) have experimented with a distance measure defined by

$$d_{ij}^2 = \left\{\frac{1 - \rho(\mathbf{x}_i, \mathbf{v}_j)}{1 + \rho(\mathbf{x}_i, \mathbf{v}_j)}\right\}^{\beta}, \tag{2.2}$$

with $\rho(\mathbf{x}_i, \mathbf{v}_j)$ as in (2.1) and $\mathbf{v}_j$ the j^{th} cluster centroid. In the implementation of EROICA, we use (2.2) with $\beta = 1$, as another choice for the distance measure. (By substituting (2.2) into (2.7), the effect of β can be combined with the fuzzy index

m, i.e., $m' = 1 + (m-1)/\beta$, hence an independent selection of β is unnecessary.) The correlation-based measure does not distinguish between identically shaped waveforms of different amplitudes.

We assessed the relative effectiveness of finding expected clusters for the two distance measures on a large number of fMRI datasets. We tested the two measures, with two normalization options, Subtract Median and Rank Order, (see the Appendix for their definition), and various combinations of the three Initial Partition methods. The Appendix also contains details of the fuzzy clustering algorithm we have developed, improved, and implemented in EvIdent®.

The spectral peak noise test is applied to all the centroids at every iteration of the FCA. The ones that fall below the noise level are removed from further analysis and the corresponding cluster member TCs are assigned to the "Rejects" cluster. This is done to eliminate ill-formed clusters since, for a well-formed cluster, the expectation is that the SP value of the centroid will be much higher than that of the average member TC. The rejected TCs are retested at Stage III against all centroids that remained above the noise level.

Stage III: Time-Course Reassignment, Significance Testing, and Cluster Membership Validation

EROICA's 3^{rd} stage involves measuring the presence of each centroid (activation profile) in *all* of the TCs in the region of interest, and assessing the statistical significance of that measure. Each TC is provisionally assigned to a cluster with whose centroid it has the largest correlation coefficient. In effect, each centroid is treated as a possible reference function (or a statistical probe), and each TC $y(n),\ n = 1, ..., T$ is modeled as a linear combination of a *single* reference function plus residual error:

$$y(n) = \beta_j x_j(n) + r_j(n) = \rho_j \frac{\sigma_y}{\sigma_{x_j}} x_j(n) + r_j(n),\ \ \rho_j = max\{\rho_l : l = 1, 2, ..., K\} \quad (2.3)$$

where j is the index of the centroid with which $y(n)$ has the largest correlation coefficient ρ_j, β_j is the model parameter (a measure of experimental effect), K is the final number of centroids produced by clustering, and $r_j(n)$ is a residual error sequence. TCs that were placed in the "trend" cluster at the initial partition stage, must be effectively detrended prior to computing ρ_j. This can be accomplished without explicit detrending by using: $\rho_j = (\rho(x_j, y) - \rho(t, u)\rho(t, x_j))/(1 - \rho^2(t, y))$, where t, x_j, y refer to the trend, reference and time-course vectors, respectively. (In this Section, we deliberately changed notation, in order to express the results in a form more familiar to practitioners of fMRI. The $\mathbf{x}_i$ and $\mathbf{v}_j$ of (2.1) are, in (2.3), y_i and x_j, respectively.)

Modeling each TC with only one reference function has the effect of partitioning the entire dataset into K groups. Therefore, on average, the number of TCs that will be tested against a particular reference function x_j is N/K, and therefore the expected number of false-positives in the final activation maps is pN/K, where p is the significance level. This amounts to a significant reduction, since K is typically

$10 \sim 20$. The first part of the significance test is to test the null hypothesis that $\rho_j = 0$. For the L TCs used in clustering, the correlation coefficient is recomputed by including a correction factor c

$$\hat{\rho_j} = \frac{\rho_j - c}{\sqrt{(1 - 2c\rho_j + c^2}} \cong \frac{\rho_j - c}{1 - c\rho_j}, \quad c \equiv \frac{\sigma_y (\mu_{ik})^m}{\sigma_{x_j} \sum_{i=1}^{L_j} (\mu_{ik})^m}$$

where L_j is the number of TCs closest to centroid j. This removes the circularity in modeling a TC with a model that is partly defined by the TC itself, and then attempting to assess the significance of ρ_j. The correction factor c removes the effect of y_i on x_j (see (2.8)); in situations where $L_j \gg 1, c$ is small and can be ignored, but when L_j is small, c can be significant. For the typically 95% of TCs that were not involved in creating x_j, this inferential circularity does not apply and no correction factor is needed.

As was the case when testing for significant trends, ρ_j in (2.3) is tested against a threshold ρ_0, whose correct value is difficult to obtain in practice because an accurate estimate of the number of effective degrees of freedom is lacking. For example, if we let $G = x_j$ and $K = I$ (the unit matrix) in the general linear model in (Friston, 1996), the test statistic becomes $t_j = \rho_j \sqrt{(T-1)/(1-\rho_j^2)}$, which is Student-$t$ distributed with $T - 1$ effective degrees of freedom. For large T, the corresponding correlation threshold ρ_0 is simply too small to be useful. To overcome this problem, we propose that ρ_0 be obtained as follows: The reference function (centroid) in (2.3) can be expressed as

$$x_j(n) = \sum_{i=1}^{m} A_{k_i} \cos(2\pi k_i n/N + \Theta_{v_k}(k)) + r_j(n)$$

where the A_{k_i} coefficients correspond to the m largest power spectrum coefficients such that

$$SP_j(k_1) \geq SP_j(k_2) \geq ... \geq SP_j(k_m) \geq SP_0(1) = -ln(p) \tag{2.4}$$

where, as was the case when detecting significant trends, $SP_0(1)$ is the lowest spectral peak threshold for the chosen level of significance p. In effect, the reference function specifies the number, locations and phases of the spectral peak coefficients at which a TC is to be tested. We take advantage of the relationship between $SP(k)$ and the correlation coefficient (see (Jarmasz and Somorjai, 2002) for a derivation) and set $\rho_{0,j} = \sqrt{SP_0(m)/M}, \quad SP_0(m) = -ln(1 - (1-p)^{\frac{1}{m}})$: the hypothesis that $\rho_0 = 0$ is rejected if $\rho_j \geq \rho_{0,j}$.

To overcome the non-specific nature of the correlation coefficient and make the significance test both more sensitive and more specific, all TCs with $\rho_j \geq \rho_{0,j}$ are also tested in the spectral domain. The test involves all m significant frequency locations in the order specified by (2.4), as established by the reference function. The null hypothesis is rejected when at least one $SP(k_i) \geq -ln[1 - (1-p)^{1/i}]$, $i = 1, 2, ..., m$. However, we allow the threshold to increase, instead of keeping it at its lowest possible value of $SP_0(1) = -ln(p)$, because the search domain for the spectral peak is increasing: if a TC fails the first i tests, the search domain is

expanded by an additional frequency location and the test becomes $SP(k_{i+1}) \geq -ln[1-(1-p)^{1/(i+1)}]$, until $i = m$. The tests in the two domains are not independent, and overlap to some degree, but overall sensitivity and specificity are both enhanced by combining the two tests. For the model in (2.3), the spectral peak statistic is defined as (Jarmasz and Somorjai, 2002)

$$SP(k_i) \equiv \frac{|Y(k_i)|^2}{N\sigma_r^2} = \frac{SP_{IP}(k_i)}{(1-\rho_j^2)} \qquad i = 1, ...m.$$

where $SP_{IP}(k_i)$ is the spectral peak value from the Initial Partition stage and $Y(k_i)$ is the DFT coefficient. The inclusion of a reference function in the model reduces the noise variance to $\sigma_r^2 = \sigma_y^2(1-\rho_j^2)$. Consequently, the SNR as a function of frequency is given by $|Y(k)|^2/|R(k)|^2 = sin^{-2}(\sigma_y(k) - \sigma_x(k))$, i.e., the residual power at k is zero if $y(n)$ and $x_j(n)$ are in phase at k. The complete significance test with respect to a reference function $x_j(n)$ is summarized by:

$$\rho_j \geq \sqrt{\frac{-ln(1-(1-p)^{\frac{1}{m}})}{M}} \quad \text{and at least one} \quad SP(k_i) \geq -ln(1-(1-p)^{\frac{1}{i}}),$$

$i = 1, 2, ..., m.$

The two-domain significance test, together with clustering, eliminates most of the expected false-positive time-series from the activation maps.

EROICA's third stage Significance Test is statistically superior to previous approaches, because, for a given significance level, it uses a better estimate of the true number of statistically significant degrees of freedom in the activation-profile. In addition, the test extracts a statistic with a well-defined probability distribution, which yields an accurate threshold. This third stage not only updates the clusters originally identified at Stage II (FCA), by allocating some of the initially excluded TCs to them, it also "purifies" and "validates" these clusters by removing TCs that fail specific statistical criteria. Consequently, the number of "interesting" TCs initially found at Stage II will likely change (increase or decrease) at the completion of the full three-stage analysis. To assess how coherent a cluster is, we have suggested (Baumgartner et al., 1999) computing $W (0 \leq W \leq 1)$, Kendall's measure of concordance (Kendall and Gibbons, 1990). W tests, in a pairwise fashion, the inherent similarity amongst the TCs comprising the cluster, without comparison with some reference, such as the cluster centroid or an externally defined reference function. ($W \approx 0$ for a group of noise TCs; W = 1 if all TCs are identical in their rank order. W is essentially a scaled average of all pairwise Spearman correlation coefficients of the TCs comprising the cluster.)

2.2.1 The Datasets

2.2.1.1 "Phantom" Datasets

We have analyzed four phantom datasets. These were created from a real null fMRI experiment (i.e., the subject in the magnet was at rest, not presented with any task). The single-slice single-shot EPI dataset, acquired on a GE Signa scanner ($FA/TE = 90^o/50$, matrix size 128×128, 120 scans, TR = 3500 ms), was thresholded to 3975 brain pixels. A 46-pixel contiguous region was excised, and the excised pixel TCs were replaced with a two-peak "activation," each peak having been constructed from a two-parameter gamma function simulating the hemodynamic response (Lange, 1996). Gaussian noise was added to each of these activated TCs, such that the contrast-to-noise (CNR) values were 2, 3, 4 and 5, i.e., realistic values for typical fMRI experiments ($CNR = \Delta S/\sigma_{noise}$, where ΔS is the signal enhancement and σ_{noise} is the noise standard deviation (Lange, 1996)). For the default normalization option (SM), the analyses were carried out with both Euclidean (E) and correlation-based (C) distance measures. We tested all three Initial Partition options in different combinations; For the Trend Exclusion (TE), Auto-Correlation (AC), and/or Spectral Peak (SP) we set $p = 0.05$. For the allocation threshold, $p = 0.01$. We set the Merging Index to 2 (see the Appendix).

2.2.1.2 Studies of Execution Times

Sixteen fMRI datasets were analyzed to assess the effect of Initial Partition on the quality of the clustering results, and on computational speed. To evaluate the robustness of the results, we used six normalization options in EvIdent® (see the Appendix), and the two distance measures (Euclidean and correlation-based). Thus, we completed 12 analyses for each dataset, 192 analyses in total. For all these analyses the trend exclusion ($p = 0.01$), autocorrelation ($p = 0.01$) and merging (Index = 4) options were enabled. The datasets included both simple and more complex motoric, visual, and cognitive/motoric paradigms. One of the datasets derives from a mental rotation task (Richter et al., 2000), presented to the same subject at sixteen separate occasions; we concatenated these 16 T-scan experiments into a composite 16T-scan dataset. Some datasets have clinical relevance, including one for a patient with Tourettes syndrome (Gates et al., 2001), and another with self-monitored onset and cessation of seizures.

2.2.1.3 Analysis of Real Data

We report full analyses for one of the more complex datasets. (Additional results, for both a smaller and a larger dataset are given in (Somorjai et al., 2001).) This is an 8-slice, 50-scan set ($N = 14,957,128 \times 128$), involving a finger-tapping paradigm, executed first with the left, then with the right hand. This dataset demonstrates the importance of the three stages, and the robustness of the results, once the Initial

Partition options were enabled. We show results only for the default normalization option, Subtract Median, and the correlation-based (C) distance measure. We display and discuss the results for several combinations of Initial Partition options. We selected $p = 0.01$ both for the TE + AC and TE + SP options. The p-value for TC reassignments (Significance Test, Stage III) was also set to 0.01. Although computations were carried out both with merging "off," and with merging "on," for brevity we report results only for the latter, with the Merging Index at its default value, 4.0. All analyses were carried out with our software implementation of the three-stage process EROICA. The software is called EvIdent® (EVent IDENTification) (www.ibd.nrc.ca/informatics), and it also contains basic image processing functions. The analyses of the execution times were carried out with EvIdent 5.0, compiled and running on an SGI Origin 2000, 180 MHz R10000 CPU (this has approximately the same execution speed as a 350 MHz Pentium II PC). Unless otherwise indicated, we kept the following parameters at their default values: fuzzy index $m = 1.1$, initial number of clusters = 35, maximum number of iterations = 25, and centroid initialization = Maximum Dispersion (this uses the deterministic and reproducible maximin distance algorithm (Tou and Gonzales, 1974)). In no case did FCA surpass the maximum number of iteration allowed.

2.3 Results

2.3.1 Phantom Studies

The phantom studies were designed to answer the question: How successful is FCA in identifying activation in the presence of different levels of noise contamination, and what are the advantages if any, of the various Initial Partition options?

We show the results in Table 2.1 for CNR = 2 and 3. Inspection of the Table suggests that for CNR = 2.0, the number of FPs (false positives) and/or FNs (false negatives) do not depend strongly on which or how many of the Initial Partition options were enabled, as long as at least one of them was. For CNR = 3.0, the FPs are 0 for seven of the eight combinations. In one case FP = 1. In all cases FN = 0.

For CNR = 2.0, the median of Kendall's concordance values over the eight combinations is $W_{median} = 0.287$, whereas the median of the corresponding average Pearson correlation coefficients between the cluster centroids and the TCs in the clusters is $\rho_{ave}^{median} = 0.556$. For CNR = 3.0, $W_{median} = 0.410$, $\rho_{ave}^{median} = 0.704$. For both CNR = 4.0 and 5.0 (not shown in the Table 2.1), FP = FN = 0 in all cases. $W_{median} = 0.478$, $\rho_{ave}^{median} = 0.792$ ($CNR = 4.0$), $W_{median} = 0.519$, $\rho_{ave}^{median} = 0.849$ ($CNR = 5.0$). The results are similar when using the Rank Order normalization option.

Table 2.1 "Phantom" Datasets

	Initial Partition Options			CNR = 2.0				CNR = 3.0			
	TE	SP	AC	FP : FN	W	ρ_{ave}	Speedup ratio	FP : FN	W	ρ_{ave}	Speedup ratio
E – Measure	N	N	N	4 : 4	0.294	0.568	(1.0)	0 : 0	0.410	0.707	(1.0)
	Y	N	N	1 : 3	0.298	0.567	2.1	0 : 0	0.410	0.704	1.9
	Y	Y	N	4 : 0	0.280	0.549	7.2	0 : 0	0.410	0.704	9.5
	Y	N	Y	1 : 0	0.290	0.555	12.6	0 : 0	0.410	0.704	12.6
C – Measure	N	N	N	2 : 2	0.290	0.567	(1.0)	0 : 0	0.410	0.705	(1.0)
	Y	N	N	3 : 0	0.283	0.556	2.1	1 : 0	0.402	0.696	3.0
	Y	Y	N	3 : 0	0.281	0.552	7.4	0 : 0	0.410	0.704	10.3
	Y	N	Y	3 : 0	0.279	0.546	11.0	0 : 0	0.410	0.704	15.6

Results for "phantom" datasets obtained by choosing different combinations of Initial Partition options, either for Euclidean (E) or for correlation-based (C) distance measures. The correct number of artificial "activated" TCs is 46. FP = false positive, FN = false negative. Initial Partition options: TE = trend exclusion, SP = spectral peak, AC = autocorrelation. Y: option enabled; N: option disabled. W: Kendalls measure of concordance ($0 \leq W \leq 1$). ρ_{ave}: average value of Pearson's correlation coefficient between the cluster centroid and the individual TCs of the cluster. The speedup ratios are relative to the execution times when all three Initial partition options are disabled.

2.3.2 Studies of Execution Times

To assess how different combinations of the available options influence execution times, we have analyzed sixteen fMRI datasets of widely different types and sizes, for all combinations of the six normalization and three Initial Partition options, and the two similarity measures, using the aforementioned default values and rules. The number N of brain voxels analyzed ranged from 3500 to 68000, the number T of scans from 35 to 1408, a wide range of realistic N and T combinations.

The FC execution times t_{exe} spanned 0.4 to 29 seconds, and did not depend strongly on which of the six normalization options were selected (the t_{exe}s were within 10–15%). Based on these 16 datasets, we fitted the t_{exe}s (averaged over the various combinations of options) to $N^{\alpha}\ T^{\beta}$, where N is the total number of brain voxel TCs and T is the number of scans.

The 3-parameter expression $t_{exe} = AN^{\alpha}T^{\beta}$, with $A = 9.7118 \times 10^{-6}, \alpha = 0.8912, \beta = 0.9131$ gives a good fit, with the coefficient of determination $R^2 = 0.9689$. (An easy-to-remember upper bound approximation is $t_{exe} < BNT$, i.e., execution times are approximately linear in both the number of brain TCs and the number of time instances. $B = 10^{-5}$ for the 180 MHz Intel CPU.) The fraction of

the N original TCs chosen (i.e., passing either trend exclusion and autocorrelation tests (TE + AC) or trend exclusion and spectral peak tests (TE + SP)) is obviously strongly data-dependent. For the sixteen fMRI datasets, the median percentage of preselected TCs was 4.2%. The median speedup ratios relative to no Initial Partition were, for (E, AC) = 4.8, (E, SP) = 4.7, (C, AC) = 5.9 and (C, SP) = 4.3, where E indicates using the Euclidean metric, C the correlation metric. The major differences are between the TE "off" and TE "on" groups. For the former, the average speedup is $\sim$ 2-fold, for the latter $\sim$five-fold.

2.3.3 Analysis of Real Data

The task paradigm consisted of one activation epoch of finger tapping, first by the left, and then by the right hand. Ideally, one would expect corresponding

Table 2.2 Real Dataset

	Finger-Tapping, Left Hand			
Initial Partition	**None**	**TE**	**TE + SP**	**TE + AC**
Number of TCs; (ρ_{ave})	-	434; (0.607)	159; (0.651)	216; (0.663)
	Finger-Tapping, Right Hand			
Initial Partition	**None**	**TE**	**TE + SP**	**TE + AC**
Number of TCs; (ρ_{ave})	-	-	159; (0.625)	150; (0.627)
No. of final Trend TCs	-	5224	4997	5082
No. of final Rejects TCs	9251	7768	7155	7567
No. of TCs Clustered On (Percent of Total)	14957 (100.0)	9036 (60.4)	433 (2.9)	1337 (8.9)
t_{exe} (Relative)	(1.00)	0.71	0.11	0.20

Results of the analyses of an 8-slice, 50-volume dataset (courtesy, Prof. E. Moser). Paradigm: one activation epoch each of finger tapping, left hand first, then right hand. The columns correspond to four Initial Partition (IP) options: IP disabled (None), TE enabled, TE + SP enabled, TE + AC enabled. The first row contains the number of TCs in the left-hand finger-tapping cluster, and their average Pearson's correlation coefficient (ρ_{ave}) with the centroid. The 2^{nd} row lists the same quantities, for the right-hand finger-tapping cluster. The 3^{rd} shows the final number of TCs assigned to the "trend" cluster, the 4^{th} the ones assigned to the "rejects" cluster. The ^{5th} row shows the number of TCs to which FC was applied, with the percentage of the total number in parentheses. The last row lists the execution times relative to the "No IP" option.

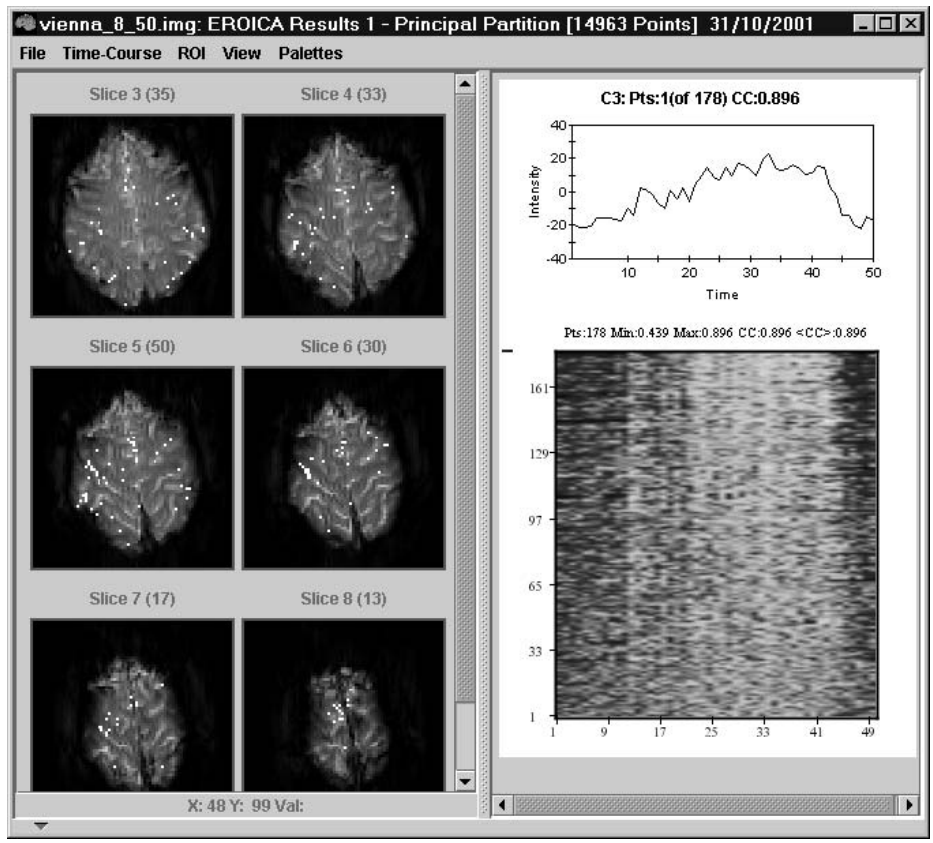


Figure 2.2 Homogeneity maps, centroids and activation maps for an 8-slice (only slices 3–8 are shown), 50-scan fMRI dataset, using FCA with SM normalization, C- metric, Merging Index = 4.0: cluster for the right hand finger-tapping activation.

responses in the appropriate right and left motor areas. This is a particularly difficult dataset to analyze because many of the low frequency confounds due to aliasing are indistinguishable from the actual activation TCs. In Table 2.2, we list the outcomes of using several combinations of the Initial Partition options, including clustering on the entire 14,957-TC set (i.e., no Initial Partition, NNN). Without any Initial Partition, no activation could be cleanly recovered. We have verified that this was not because of possibly excessive cluster merging (with the default merging index MI = 4.0, the merging algorithm merged the original 35 initial clusters into 17). Even without merging, and starting with 50 initial clusters, the responses were swamped by "noise" and "trend" TCs.

In fig. 2.2, we display the two expected activation maps for brain slices 3–8. The cluster centroids correctly reflect the finger-tapping paradigm executed by the right hand (fig. 2.2) and by the left hand (fig. 2.3). The homogeneity map (Baumgartner et al., 1999) displayed beneath the centroids depict the TCs on the given slices (vertical axis) vs. Image Number (time instance, horizontal axis). These TCs are ordered: the one on top correlates the most strongly with the centroid, the one at

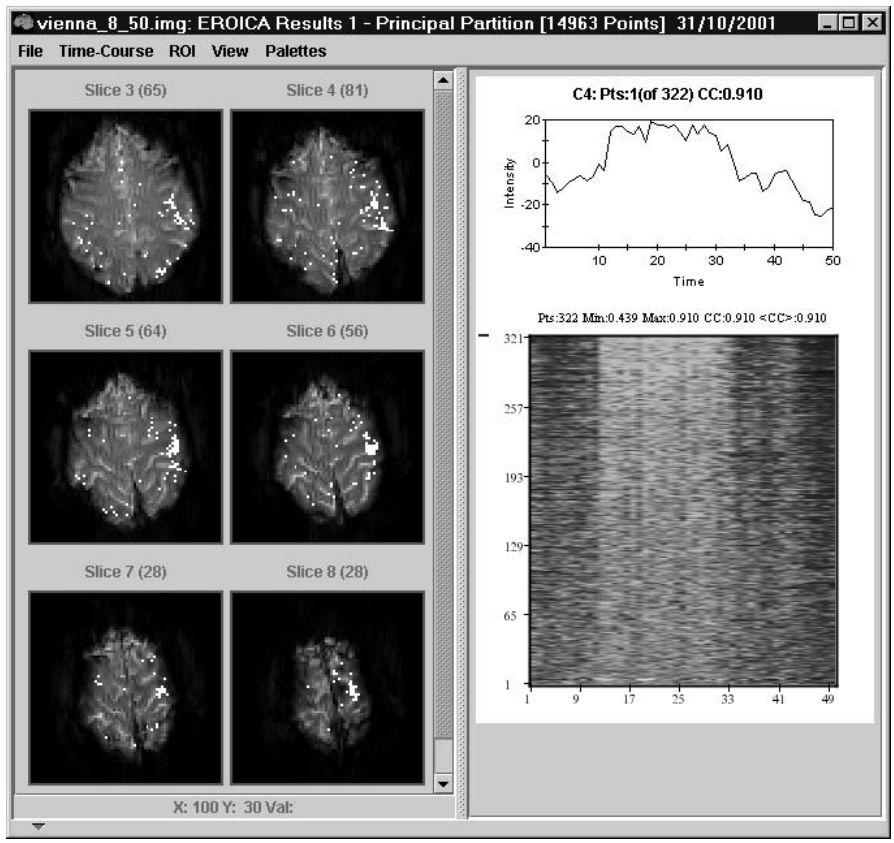


Figure 2.3 Homogeneity maps, centroids and activation maps for an 8-slice (only slices 3–8 are shown), 50-scan fMRI dataset, using FCA with SM normalization, C- metric, Merging Index = 4.0: cluster for the left hand activation.

the bottom least strongly. The homogeneity map provides a useful visual display of the internal consistency of the TCs comprising the cluster (also summarized, as a single number, by ρ_{ave} or Kendall's concordance W). As shown in Table 2.2, selecting only trend exclusion (with $p(\text{TE}) = 0.01$) wasn't sufficient to identify the right-hand finger tapping activation profile; enabling TE alone did pick out the left-hand activation region (see Table 2.2 for more details). We find the two clusters of interest when both TE and SP ($p = 0.01$) or both TE and AC ($p = 0.01$) are selected. Closer inspection of Table 2.2 reveals that selecting only TE reduced the number of TCs to be considered at the Principal Partition stage to 60.4% (9036 TCs) of the original 14,957 TCs. The relative execution time t_{exe} dropped to 71% of the full computation time. When both TE and AC are selected, only 8.9% (1337 TCs) are used by FCA, and t_{exe} is down to only 20%. Finally, when both TE and SP are enabled, only 2.9% (433 TCs) are used by FCA, and t_{exe} drops further to 11% of the full analysis.

2.4 Discussion

2.4.1 Phantom Studies

Results in Table 2.1 suggest that, at least for these phantom datasets, the major advantage is execution speedup. For both the Subtract Median and Rank Order (results not shown) normalization options, and for both E- and C-measures, the artificially created activation was recovered, even when the "No Initial Partition" option was chosen. The quality of the recovered clusters, as measured both by W and by ρ_{ave}, is comparable for "No Initial Partition" and for the other three combinations of Initial Partition options. The main distinction is that for CNR = 2.0, the "No Initial Partition" option produced more false positives and negatives, and in general, the results with the C-measure were somewhat less accurate than with the E-metric. For CNR = 3.0, there are no FNs and only one of the eight combinations produced a single FP. For CNR values larger than 3.0, recovery of the "activated" TCs is error-free in all cases (not shown). Thus, an approximately 2-fold speedup was obtained with the TE option enabled; the speedups increased to approximately 12-fold with TE and one of the two Initial Partition options (AC or SP) enabled, all producing comparable or better cluster recovery than the "No Initial Partition" option. As expected, with increasing CNR the centroids become progressively less noise-corrupted, reflecting the increased homogeneity (larger Ws) of the activated cluster. Furthermore, as CNR increases, the TCs belonging to the cluster correlate better (larger values of ρ_{ave}) with their cluster centroid.

2.4.2 Studies of Execution Times

The execution time results indicate that there is no significant statistical difference between the FC analyses carried out with the two distance measures. The differences are more pronounced among the six normalization options (not shown). Subtract Median and Rank Order appear to be the most reliable normalization options in terms of overall clustering success (i.e., clean recovery of the response to the task paradigm). McKeown et al. (1998) report 90 minutes execution time on a DEC Alpha 2100A workstation for the spatial ICA method, on an fMRI dataset comprising 135–146 images and 8–10 slices. Estimating that their data comprises ∼25,000 brain voxels, for a comparable-sized dataset (N = 26,293, 21 slices, 140 time instances), with Initial Partition enabled, the execution time with EvIdent® is $\sim 5-6$ secs. The current flurry of activity in the literature (McKeown, 2000; Carew et al., 2001; Laconte et al., 2001; Lin et al., 2001; Rogers et al., 2001; Calhoun et al., 2001) applying ICA to fMRI data all suggest that extensive preprocessing is needed to obtain acceptable execution times. The standard approach is to first use PCA to reduce the number of components to be extracted by ICA to a computationally tolerable level (typically 50–80 PCs) (McKeown, 2000).

The correlation-ordered homogeneity maps of figures 2.2, 2.3 provide a quick and ready visual check of the internal consistency of the TCs belonging to the parent cluster. Because cluster centroids are (membership-weighted) averages of the constituent TCs, they are naturally less noisy than the individual TCs and hence reveal the temporal behavior more clearly. They provide an intrinsic, data-determined model of the brain's response; they can be used as natural inputs for subsequent inferential tests.

2.4.3 General Considerations

It should be emphasized that we analyzed the majority of the 16 datasets blindly, without knowing anything about the task paradigm. (We invariably found the "correct" (i.e., expected) response that was communicated to us *post facto* by the original investigators.) Thus, the conditions were challenging: in normal circumstances the investigator who designs the experiment, would have some idea of what to expect. However, we wanted to simulate the conditions of a clinical study or a complicated cognitive/linguistic experiment, for which such knowledge may not be available.

The importance of always enabling the Trend Exclusion (TE) Initial Partition option, (and one or both of the Autocorrelation (AC) or Spectral Peak (SP) options) is incontestable. Because of the Initial Partition, the actual clustering is, on average, done only on $\sim 4\%$ of the voxels. This is part of the reason for the much enhanced execution speed of EROICA. (Some of the other reasons are technical and are detailed in the Appendix.) Furthermore, it is much more likely that "interesting" clusters are obtained if most of the confounding effects of "trends" and "noise" are initially excluded. The Significance Test (Stage III) guarantees that small effects, initially excluded, are not missed. Alternatively, the initially excluded TCs can be re-analyzed. Whether we select (along with TE), AC, SP or both, depends on the data. If the duration of the activation is short, then AC may not always work, especially for low SNR. If the SNR is low and the paradigm is "busy," i.e., there are many (quasi) periods, then SP tends to be more effective.

When prior knowledge is available about the expected response, performing the t-test, or computing the correlation coefficients between the brain TCs and a predefined reference function could and should be viewed as a type of initial partition of the TCs. We would still apply the FCA method (or any equivalent EDA approach) to let the data define the precise shape of the response, and eliminate many false positives, before any inferential method is used (McIntyre et al., 1998).

The three-stage strategy we have described is beneficial for any EDA method, whether it uses FCA (as does EROICA) or some other unsupervised pattern recognition method. EROICA satisfies the basic requirements of a viable, efficient EDA method for analyzing fMRI data. Based on more than fifty fMRI datasets, acquired with both FLASH and EPI, on MR imaging systems with field strengths ranging from 1.5T to 4.0T, EROICA has invariably detected the "true" response, even under very noisy conditions. (Here "true" refers to the assessment of the

original designers of the various fMRI experiments that were analyzed blindly by the EROICA process.) Of course, some experimentation with parameters and p-values may be needed, but this is precisely in the spirit of EDA. We have reported (Tuor et al, 2000) a good example of the type of novelty that could be detected by an EDA process. We used EvIdent/EROICA to analyze the outcome of fMRI experiments on rats responding to noxious electrical and chemical stimuli. EROICA identified several clusters of TCs, whose centroids graphically depict an unexpected, *sequential progression* of activations to pain, occurring first in either the right or the left hemisphere, with a separation of seconds to minutes between peak activations. The result of pre-treatment with morphine was also clearly identified: activation response to electrical stimulation was inhibited in most regions except for sensory-motor cortex.

One might argue that the "potentially interesting" TC centroids (testable hypotheses) found ought not be tested against the very dataset from which they were derived. The three-stage strategy practically eliminates this statistically undesirable circularity. In fact, the Stage I initial partition (preselection) eliminates the majority of the TCs prior to FCA. The newly found hypotheses, generated at the second stage from a small fraction of the total number of TCs, can be tested with reasonable statistical confidence against the remainder of the data (or, fully legitimately, against data obtained from repeated experiments on the same subject or on data produced by the same experiments but on different subjects. As already discussed, we would still submit the TCs so identified to Stages II and III, to discover subtle inter-experiment or inter-subject differences.). We accomplish this post-analysis testing with the conservative allocation and clean-up procedure of Stage III. However, allocation may lower the overall intrinsic homogeneity of an enlarged cluster. Because of this possibility, as a useful cluster validation strategy, we propose to "purify" the clusters in some fashion, e.g. by the algorithms described in (Baumgartner et al., 1999; Baumgartner et al., 2001), (Davison and Somorjai, 2001). (Alternatively, EROICA can be repeated with a more stringent p-value, i.e. larger ρ_{min}.)

It is important to emphasize that an EDA method, such as FCA or ICA, is only useful in practice if a large number of different analyses can be conducted in computationally reasonable times, e.g., at most a few minutes per dataset for a complete analysis (Requirement 3, speed). Examples of what can be accomplished with a fast EDA method include the possibility of analyzing experiments over several subjects for common responses (e.g. by stacking the K brain slices of each of M subjects to create a KM slice "meta brain" (Somorjai, 2001), or concatenating the TCs by combining L experiments of T scans each into a composite LT-scan dataset ("meta experiment") (Richter et al., 2000), (Somorjai, 2001).

fMRI data are sufficiently varied and complex, so that it is unlikely, and unreasonable to expect, that there are theoretical criteria by which, for any given dataset, an optimal preprocessing option/distance measure could be selected a priori. What we do propose is a data analysis strategy that is useful, or even essential, for complicated cognitive/linguistic tasks or clinical applications. We do not believe that

an automated "black box" approach with pre-set recipes is desirable or even feasible for large, complex data. There is no best prescription, especially since what is deemed best depends on the question one wants answered. This view of using several different analysis methods is gaining acceptance in the fMRI data analysis literature (Lange et al., 1999). Despite these caveats, we can suggest certain guidelines for analyzing fMRI datasets. Some of these are generic, applicable to any EDA method; others are specific to using FCA for the second stage.

Given the size and complexity of fMRI datasets, identification and temporary exclusion of trend and noise TCs is critical for finding the relatively small number of potentially interesting TCs at the first, initial partition stage. Once these interesting TCs have been grouped by some means (e.g. by FCA), the initially excluded TCs are to be assigned to the groups, and the ultimate groups' homogeneity (internal coherence) checked and validated (stage three). If the actual group/cluster identification is to be performed via fuzzy clustering, it is important to start the process with a large number of initial cluster "seeds." This ensures that the high-dimensional T-space of the TCs is adequately sampled.

We have repeatedly emphasized that a viable EDA method must have the flexibility, versatility, and robustness/reliability to allow a systematic and comprehensive, yet acceptably fast exploration of the data. We have shown that EROICA, the proposed three-stage process, with appropriate TC preselection, followed by FCA, and completed by an appropriate significance test (i.e. the possibility of reassigning initially excluded TCs) and cluster validation, readily satisfies the conditions that qualify it as a viable EDA method.

Of course, we are not suggesting that the EDA method used for analyzing fMRI data must necessarily be based on FCA. In particular, spatial ICA seems to be gaining some acceptance as a possible EDA method for fMRI. However, we do not regard ICA as a direct competitor of FCA (just as FCA is not meant to supplant hypothesis-driven inferential methods). At least for now, ICA still lacks the flexibility, versatility and speed that a competitive EDA must have. (Frequently, even after PCA-based feature reduction, some hybrid version (e.g., HYBICA (McKeown, 2000)), which combines the data-driven ICA with a priori hypothesis-guided methods, needs to be used to complete the analysis.) We view ICA as complementary to FCA. Although both unsupervised, they address different aspects of the general data-analytic problem. ICA could be used in subsequent analyses, if it appears that the centroids are (linear) mixtures of well-defined temporal shapes. That this view is reasonable is supported by (Karhunen and Malaroiu, 1999). These authors, although strong proponents of ICA, realized its limitations (linearity, description of data by the same global features), and proposed to first "preprocess" the data by k-means clustering, and then analyze / demix each of the K clusters (or their centroids) by local ICA.

The most important conceptual feature of an EDA method is that it is data-driven, and thus bias- and model-free. Its primary role is to help generate new hypotheses (i.e., models) directly from the data, to be subsequently tested and verified by some model-driven inferential method. Ideally, an EDA method should

precede the more conventional statistical inferential methods of analysis (Somorjai and Jarmasz, 1999), i.e., complement, not supplant the latter. This is what EROICA was designed to do. In practice, exploration-confirmation-inference is likely to be carried out iteratively. In a landmark paper (Tukey, 1962) argued that classical statistics, with its propensity for analyzing small, homogeneous, stationary, i.i.d. data, using known distributional models and assumptions, will be inadequate to handle the problems encountered in the analysis of large, complex data. The statistical community (Fayyad and Smith, 1999; Huber, 1999) increasingly endorses and verifies this prophetic view. As Huber succinctly stated in his classic review on Projection Pursuit (Huber, 1985): "There are no panaceas in data analysis." The entire concept of EDA is an acknowledgement of this fact.

2.5 Conclusion

We have argued that the analysis of the typically large, complex, heterogeneous fMRI data ought to start with an EDA method that allows the data to reveal its inherent structure without prior models or assumptions. EROICA, the three-stage strategy of preprocessing / analysis / inferential validation we propose for fMRI, at its first stage identifies, by various direct and indirect noise tests (preselection methods), "potentially interesting" temporal structure amongst the TCs, and at the second stage, focuses analysis on these via an appropriate unsupervised pattern recognition method. At the third stage, we attempt to allocate the initially excluded "trend" and "noise" TCs to the clusters they most resemble, and validate and "clean up" the second-stage clusters by possibly reassigning its original members. The first stage renders large datasets computationally manageable, a critical requirement. At the second stage, our choice of fuzzy clustering over ICA or PCA/FA is partly dictated by speed of execution, but more importantly, by the heterogeneous nature of large fMRI data: FCA is local, (i.e., the cluster centroids do not consist of linear combinations of all voxel TCs) hence does not get confounded by global heterogeneity. The third stage is aimed to assure that the hypotheses (i.e. the ultimate cluster centroids) generated by the EROICA process are minimally contaminated by the influence of false positive TCs. Overall, the three stages of EROICA fulfill the requirements for a practically feasible and useful exploratory analysis of fMRI data.

Acknowledgments

Over the last six years, a number of people contributed to our software EvIdent®. We thank Gord Scarth who helped develop and implement the early, IDL-based version of EvIdent, Dick Jackson and Walter Roberson, who improved and generalized the IDL user interface, Richard Baumgartner for creating the artificial datasets from a null set, and Lawrence Ryner who provided it, Nick Pizzi and Rod Vivanco for codesigning the Java-based software system, and the latter for implementing it.

The following colleagues provided fMRI datasets: Dr. Stephan Bucher (3 sets), Dr. Larry Gates (2 sets), Prof. Michael McIntyre (2 sets), Prof. Ewald Moser (3 sets), Dr. Wolfgang Richter (3 sets) and Prof. Keith Thulborn (3 sets).

2.6 Appendix

2.6.1 Preprocessing Details

Aside from not preprocessing the raw TCs prior to FCA (leading to segmentation), the normalization options in EvIdent® are Subtract Mean, Subtract Median, Divide by Median, Robust (defined as "subtract median and divide by MAD," the Median Absolute Deviation), Non-Robust (defined as "subtract mean and divide by standard deviation"), and Rank Order. Some of these eliminate from the TCs not only the level dependence but also the influence of different variances (or their robust equivalent, MAD). In particular, by Rank Orderingrank ordering the T intensity values of each TC (i.e., replacing them by their rank from 1 to T), and scaling them to the $(1/T, 1)$ range by dividing each by T, both the mean $(= (T-1)/2T \approx 0.5$ for large $T)$ and the variance $(= (T+1)/12T \approx 1/12$ for large $T)$ will only depend (weakly) on T, and be the same for all TCs. (In fact, by Rank Ordering, all moments, not only the first two above, become the same for all TCs; they all depend only on T.) Two other important characteristics of Rank Ordering are that 1) it monotonically and nonlinearly transforms (any nondecreasing function $g(I_k)$ of) the original intensity values I_k, i.e., $R(g(I_k)) = R(I_k)$, and if $g(I_k) > g(I_m)$ then $R(I_k) > R(I_m)$, where $R()$ is the rank; 2) whatever the original distribution of the T intensities, the rank-ordered ones are uniformly distributed. The first characteristic eliminates any undue influence from outliers (excessively large intensities).

If the "subtract" option is chosen prior to FCA, then FCA creates clusters that differ in their TCs' absolute activation amplitude above a common (zero) resting state. This is the desired outcome if we wish to distinguish between apparent activations in large vessels and small activations in the cortex (Baumgartner et al., 1997). If the "divide" option is selected, then FCA produces clusters comprised of TCs with a common baseline (set to unity), but with different relative (%) changes in the activations above this baseline. This is how results are most frequently reported in the fMRI literature. Note that when Subtract Mean is used for normalization, $d_{ij}^2 = tan^2(\theta_{ij}/2)$, with θ_{ij} the angle between the vectors $\mathbf{x}_i$ and $\mathbf{v}_j$.

2.6.2 Temporal Fuzzy Clustering Methodology

Hard k-means *temporal* clustering of brain fMRI data was proposed in (Ding et al., 1994) and further investigated, together with hierarchical clustering, in (Goutte et al., 1999). A recent hard k-means-based version that includes cluster merging is reported in (Baune et al., 1999). A hybrid of hierarchical and k-means clustering was

suggested recently (Filzmoser et al., 1999) and applied to fMRI. Independently, in 1995 we introduced to fMRI analysis the fuzzy c-means variant (Scarth et al., 1995; Scarth et al., 1996) which, based on our previous experience (Gordon and Somorjai, 1992), seems to have a definite advantage over the hard version. In particular, unlike its hard counterpart, the fuzzy clustering algorithm is much less prone to converge prematurely to an unsatisfactory local minimum (or critical point) of the objective function that is being minimized to achieve cluster separation. This has been confirmed in another application (Geva and Kerem, 1998). Given N voxel time-courses (TCs) of T time points (scans) each, partitioned among an a priori fixed number of K clusters, the original Bezdek algorithm (Bezdek, 1981) minimizes the objective function $J_m(U,V)$

$$J_m(U,V) = \sum_{n=1}^{N}\sum_{k=1}^{K}(u_{kn})^m d_{kn}^2 \tag{2.5}$$

$U = \{u_{kn}\}$ is the $K \times N$ matrix of fuzzy membership values, $m \geq 1$ is the fuzzy index. The distance d_{kn} is (typically but not necessarily) the Euclidean distance from the n^{th} TC $\mathbf{x}_n = \{x_{n1}, x_{n2}, x_{n3}, ..., x_{nT}\}$ to the k^{th} cluster centroid $\mathbf{v}_k = \{v_{k1}, v_{k2}, v_{k3}, ..., v_{nT}\}$

$$d_{kn}^2 = \sum_{t=1}^{T}(x_{nt} - v_{kt})^2, \quad 1 \leq n \leq N, \quad 1 \leq k \leq K. \tag{2.6}$$

The minimization is via a two-stage iterative process (Picard iteration). After initializing the K cluster centroids, first one calculates the distances d_{kn}^2, followed by cluster membership values $\{u_{kn}\}$, $1 \leq n \leq N, 1 \leq k \leq K$, from

$$u_{kn} = (\sum_{j=1}^{K}(d_{kn}/d_{jn})^{2/(m-1)})^{-1}. \tag{2.7}$$

The second stage consists of updating the cluster centroids $\mathbf{v}_k$, using:

$$\mathbf{v}_k = (\sum_{n=1}^{N}(u_{kn})^m \mathbf{x}_n)(\sum_{n=1}^{N}(u_{kn})^m)^{-1}, 1 \leq k \leq K. \tag{2.8}$$

Equations (2.6), (2.7) and (2.8) are iterated alternately until J_m of (2.5) converges. The FC variant that we have developed and implemented in our EDA software EvIdent® is a significant reformulation of the classical Bezdek algorithm, expressly optimized for computational efficiency. The improvements are at both the coding and algorithmic levels. For example, we have determined empirically that, in (2.8), TCs need contribute only to the two closest centroids. This does not adversely affect accuracy, as long as TC membership values are sufficiently high. Also, distances to centroids that have changed little from the previous iteration need not be recomputed. These measures have led to a 2–3-fold speedup relative to the classical algorithm. We have also found that finding good solutions is generally more likely if the cluster centroids are initialized, instead of the membership matrix. Furthermore, centroid initialization by a reformulated version of the deterministic

Maximum Dispersion minimax initialization algorithm (Tou and Gonzales, 1974) generally works better than random initialization. The success of clustering is further enhanced by using as initial seeds some of the SP-identified centroids.

After some earlier experimentation with different distance measures, including the Mahalonobis distance, we chose either the Euclidean or the correlation-based distance measure when computing the distance between a voxel time-course and any of the K cluster centroids. This resulted in significant execution speedup over the Mahalonobis distance, without any noticeable disadvantage (our distance measures do not consider or require computing the covariance matrix). (In fact, the Mahalonobis distance tends to produce "fuzzier," less well-defined clusters.) We have introduced further conceptual advances and algorithmic improvements to the FCA. In particular, we have developed a novel cluster- merging algorithm (an earlier version was announced in (Jarmasz and Somorjai, 1998)).

Cluster merging (Jarmasz and Somorjai, 1998) is an additional option in EvIdent® that helps speed up the execution of FCA. The algorithm considers only those pairs of clusters for merging that are mutually closest to each other. It uses both centroid proximity (similarity) and membership-based criteria. The degree of merging is controlled by a user-selectable parameter. In EvIdent, this parameter ranges between 0 (no merging) and 10 (maximal merging). Combined with the Initial Partition methods discussed in the main body of the MS, a 50–90-fold speedup over the older, fixed-K versions (see also Moser et al., 1999) was achievable.

References

W. Backfrieder, R. Baumgartner, M. Samal, E. Moser, and H. Bergmann. Quantification of intensity variations in functional MR images using rotated principal components. Phys.Med.Biol. 41, 1425-1438, 1996.

M. Barth, J.R. Reichenbach, R. Venkatesan, E. Moser, and E.H. Haacke. High-resolution, multiple gradient-echo functional MRI at 1.5 T. Magn Reson Imaging 17, 321-329, 1999.

R. Baumgartner, R.L. Somorjai, R. Summers, and W. Richter. Assessment of cluster homogeneity in fMRI data using Kendall's coefficient of concordance. J. Magn Reson Imaging 17, 1525-1532, 1999.

R. Baumgartner, L. Ryner, W. Richter, R. Summers, M. Jarmasz, and R. L. Somorjai. Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis. J. Magn Reson Imaging 18, 89-94, 2000.

R. Baumgartner, G. Scarth, C. Teichtmeister, R. L. Somorjai, and E. Moser. Fuzzy clustering improves reproducibility of gradient echo functional MRI by separating various contributions from venous vessels in the human visual cortex. Fifth Scientific Meeting ISMRM, Vancouver, B.C., Canada, 1662, 1997.

R. Baumgartner, G. Scarth, C. Teichtmeister, R. L. Somorjai, and E. Moser. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part I: Reproducibility. J. Magn Reson Imaging 7, 1094-1101, 1997.

R. Baumgartner, R. Somorjai, R. Summers, and W. Richter. Ranking fMRI time courses by minimum spanning trees: Assessing coactivation in fMRI. NeuroImage 13, 734-742, 2001.

A. Baune, F. T. Sommer, M. Erb, D. Wildgruber, B. Kardatzki, G. Palm, and W. Grodd. Dynamical cluster analysis of cortical fMRI activation. NeuroImage 9, 477-489, 1999.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7, 1129-1159, 1995.

J. C. Bezdek. Pattern recognition with fuzzy objective function algorithms. New York, Plenum Press, 1981.

J. C. Bezdek, J. Keller, R. Krisnapuram, and N.R. Pal. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. Kluwer Academic Publishers, Dordrecht, 1999.

V. Calhoun, J. Pekar, T. Adali, and G. D. Pearlson. fMRI of visual perception: Networks identified by SPM and independent component analysis. Proc. Intl. Soc. Mag. Reson. Med.(9), 2212. 2001.

J. D. Carew, V. M. Haughton, C. Moritz, B. Rogers, and M. E. Meyerand. Hybrid independent component analysis of fMRI data in the frequency domain. Proc. Intl. Soc. Mag. Reson. Med.(9), 1714. 2001.

P. A. Carpenter, and M. A. Just. Modeling the mind: very-high-field functional magnetic resonance imaging activation during cognition. Topics in Magnetic Resonance Imaging 10, 16-36, 1999.

K.-H. Chuang, M.-J. Chiu, C.-C. Lin, and J.-H. Chen. Model-free functional MRI analysis using Kohonen clustering neural network and fuzzy c-means. IEEE Trans Med Imaging 18, 1117-1128, 2000.

N. E. Davison, and R. L. Somorjai. Purifying Heterogeneous Clusters Using Multiple Correlation Clustering. Human Brain Mapping, Brighton, #10459. 2001.

X. Ding, J. Tkach, T. Ruggieri, P. Masaryk. Analysis of time-course functional MRI data with clustering method without use of reference signal. Second Scientific Meeting of the International Society for Magnetic Resonance in Medicine: San Francisco: ISMRM, 630. 1994.

M. J. Fadili, S. Ruan, D. Bloyet, and B. Mazoyer. A multistep unsupervised fuzzy clustering analysis of fMRI time series. Human Brain Mapping 10, 160-178, 2000.

U. M. Fayyad, and P. Smyth. Cataloging and mining massive datasets for science data analysis. J. Computational and Graphical Statistics 8, 589-610, 1999.

P. Filzmoser, R. Baumgartner, and E. Moser. A hierarchical clustering method for analyzing functional MR images. Magn Reson Imaging 17, 817-826, 1999.

H. Fischer, and J. Hennig. Neural network-based analysis of MR time series. Magn Reson Med 41, 124-131, 1999.

R. A. Fisher. Tests of significance in harmonic analysis. Proc.Roy.Soc.Ser.A 125, 54-59, 1929.

K. J. Friston, J. Phillips, D. Chawla, and C. Büchel. Nonlinear PCA: characterizing interactions between modes of brain activity. Phil Trans R Soc Lond B 355, 135-146, 2000.

K. J. Friston. Statistical parameter mapping: Ontology and current issues. J. Cereb Blood Flow Metab 15, 361-370, 1995.

K. J. Friston. Statistical Parameter Mapping and Other Analyses of Functional Imaging Data. In: Brain Mapping: The Methods., A.W.Toga and J.C.Mazziotta (eds), pp. 363-396. Academic Press, San Diego, 1996.

K. J. Friston, A. P. Holmes, K. J. Worsley, J.-B. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping 2, 189-210, 1995.

L. L. Gates, J. R. Clarke, A. Stokes, R. L. Somorjai, M. Jarmasz, R. Vandorpe, and S. M. Dursun. Neuroanatomy of coprolalia in Tourette Syndrome using functional magnetic resonance imaging. General Psychiatry (submitted)

A. B. Geva and D. H. Kerem. Forecasting generalized epileptic seizures from the EEG signal by wavelet analysis and dynamic unsupervised fuzzy clustering. IEEE Transactions on Biomedical Engineering 45, 1205-1216, 1998.

X. Golay, S. Kollias, G. Stoll, D. Meier, A. Valavanis, and P. Boesiger. A new correlation-based fuzzy logic clustering algorithm for fMRI. Magn Reson Med 40, 249-260, 1998.

H. L. Gordon and R. L. Somorjai. (1992) Fuzzy cluster analysis of molecular dynamics trajectories. Proteins 14, 249-264.

C. Goutte, P. Toft, E. Rostrup, F. Nielsen, and L.K. Hansen. On clustering fMRI time series. NeuroImage 9, 298-310, 1999.

J. C. Gower. A general coefficient of similarity and some of its properties. Biometrics 27, 957-972, 1971.

B. Horwitz. Data analysis paradigms for metabolic-flow data: Combining neural modeling and functional neuroimaging. Human Brain Mapping 2, 112-122. 1994.

P. J. Huber. Projection pursuit. Annals of Statistics 13, 435-475, 1985.

P.J. Huber. Huge data sets. Dutter, R. and Grossman, W. Compstat 1994: Proceedings, 3-13, Physica Verlag Heidelberg, 1994.

P. J. Huber. Strategy issues in data analysis. Conf. Stat. Sci. Honouring the Bicentennial of Stefano Franscini's birth, Monte Verita, Switzerland, 221-237, Birkhauser Verlag, Basel, 1997.

P. J. Huber. Massive datasets workshop: four years after. J. Computational and Graphical Statistics 8, 635-652, 1991.

J. E. Jackson. A User's Guide to Principal Components. Wiley, New York, 1991.

M. Jarmasz and R. L. Somorjai. Time to join! Cluster merging in unsupervised fuzzy clustering of functional MRI data. Sixth Annual Meeting ISMRM, Sydney, Australia. 2068. 1998.

M. Jarmasz and R. L. Somorjai. Exploring Regions Of Interest with Cluster Analysis (EROICA) Using a Spectral Peak Statistic for Selecting and Testing the Significance of fMRI Activation Time-Series. Artificial Intelligence in Medicine (in press)

J. Karhunen and S. Malaroiu. Local independent component analysis using clustering. Proc. 1st Int. Workshop on ICA and Signal Processing (ICA '99), Aussois, France. 1999.

T. Kato, P. Erhard, Y. Takayama, J. Strupp, T.H. Le, S. Ogawa, and K. Ugurbil. Human hippocampal long-term sustained response during word memory processing. NeuroReport 9, 1041-1047, 1998.

M. Kendall and J. D. Gibbons. Rank Correlation Methods. Edward Arnold, London, 1990.

S. Laconte, S.-C. Ngan, and X. Hu. Enhancing functional paradigm specific independent components with the AFRICA technique. Proc. Intl. Soc. Mag. Reson. Med.(9), 1715. 2001. Glasgow.

N. Lange. Statistical approaches to human brain mapping by functional magnetic resonance imaging. Stat Med 15, 389-428, 1996.

N. Lange, S. C. Strother, J. R. Anderson, F. A. Nielsen, A. P.Holmes, T. Kolenda, R. Savoy, and L. K. Hansen. Plurality and resemblance in fMRI data analysis. NeuroImage 10, 282-303, 1999.

N. Lange and S. L. Zeger. Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging. Appl Statist 46, 1-29, 1997.

F.-H. Lin , A. R. McIntosh, T. A. Zeffiro, J. Agnew, and J. Belliveau. Comparison of orthogonal and independent component analysis on dimension-reduced fMRI data in partial least squares framework. Proc. Intl. Soc. Mag. Reson. Med.(9), 1730, Glasgow, 2001.

J. Lin, D. Grier, and J. Cowan. Faithful representation of separable distributions. Neural Computation 9, 1305-1320, 1997.

G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. Biometrika 65, 297-303, 1978.

J. R. Mansfield, M. G. Sowa, J. R. Payette, B. Abdulrauf, M. F. Stranc, and H. H. Mantsch. Tissue viability by multispectral near infrared imaging: A fuzzy c-means clustering analysis. IEEE Trans Med Imaging 17, 1011-1018, 1998.

J. R. Mansfield, M. G. Sowa, G. B. Scarth, R. L. Somorjai, and H. H. Mantsch. Analysis of spectroscopic imaging data by fuzzy c-means clustering. Analytical Chemistry 69, 3370-3374, 1997.

J. L. Marchini and B. D. Ripley. A new statistical approach to detecting significant activation in functional MRI. NeuroImage 12, 366-380, 2000.

P. McCullagh and J. A. Nelder. Generalized Linear Models. London, 1989.

M. McIntyre, R.L. Somorjai, M. Jarmasz, U. Sboto-Frankenstein, and C. Sweetland. Combining parametric and non-parametric analyses of functional brain images. NeuroImage 7, S757. 1998.

M. J. McKeown. Detection of consistently task-related activations in fMRI data with hybrid independent component analysis. NeuroImage 11, 24-35, 2000.

M. J. McKeown, T. P. Jung, S. Makeig, G. Brown, S. S. Kindermann, T.-W. Lee, and T. J. Sejnowski. Spatially independent activity patterns in functional MRI data during the Stroop color-naming task. Proc. Natl. Acad. Sci. U.S.A. 95, 803-810, 1998.

M. J. McKeown and T. J. Sejnowski. Independent component analysis of fMRI data: Examining the assumptions. Human Brain Mapping 6, 368-372, 1998.

M. J. McKeown, S. Makeig, G. G. Brown, T.-.P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski. Analysis of fMRI data by blind separation into independent spatial components. Human Brain Mapping 6, 160-188, 1998.

E. Moser, R. Baumgartner, M. Barth, and C. Windischberger. Explorative signal processing in functional MR imaging. Int J. Imaging Syst Technol 10, 166-176, 1999.

E. Moser, M. Diemling, and R. Baumgartner. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part II: quantification. J. Magn Reson Imaging 7, 1102-1108, 1997.

E. Moser, C. Teichtmeister, and M. Diemling. Reproducibility and postprocessing of gradient-echo functional MRI to improve localization of brain activity in the human visual cortex. Magn Reson Imaging 14, 567-579, 1996.

L. Parra, G. Deco, and S. Miesbach. Statistical independence and novelty detection with information-preserving nonlinear maps. Neural Computation 8, 260-269, 1996.

M. I. Posner, S. E. Petersen, P. T. Fox, M. E. Raichle. Localization of cognitive operations in the human brain. Science 240, 1627-1631.(1988)

W. Richter, R. Somorjai, R. Summers, R. S. Menon, J. S. Gati, A. P. Georgopoulos, C. Tegeler, K. Ugurbil, and S.-G. Kim. (2000) Motor area activity during mental rotation studied by time-resolved single-trial fMRI. J. Cognitive Neuroscience 12, 310-320.

B. P. Rogers, J. D. Carew, M. A. Quigley, A. R. Laird, K. Arfanakis, C. H. Moritz, and M. E. Meyerand. Comparison of hybrid independent component analysis and simple regression of an event-related fMRI experiment. Proc. Intl. Soc. Mag. Reson. Med.(9), 2211, Glasgow, 2001.

G. Scarth, L. Bolinger, M. D. Schnall, and R. L. Somorjai. Gd-uptake imaging analysis of the breast using fuzzy clustering. Fifth Scientific Meeting of the International Society for Magnetic Resonance in Medicine, Vancouver, B.C.: ISMRM, 1052, 1997.

G. Scarth, M. McIntyre, B. Wowk, and R. L. Somorjai. Detection of novelty in functional images using fuzzy clustering. Third Scientific Meeting of the International Society for Magnetic Resonance in Medicine, Nice, France: ISMRM, 238, 1995.

G. Scarth and R. L. Somorjai. Fuzzy clustering versus principal component analysis of fMRI. Fourth Scientific Meeting of the International Society for Magnetic Resonance in Medicine, New York: ISMRM, 1782, 1996.

A. M. Smith, B. K. Lewis, U. E. Ruttimann, F. Q. Ye, T. M. Sinnwell, Y. Yang, J. H. Duyn, and J. A. Frank. Investigation of low frequency drift in fMRI signal. NeuroImage 9, 526-533, 1999.

R. L. Somorjai and M. Jarmasz. Exploratory Data Analysis of fMRI: Philosophy, strategies, tools, implementation. NeuroImage (Supplement) 9, 45, 1999.

R. L. Somorjai, B. Dolenko, G. Scarth, and R. Summers. How different are the clusters in fMRI? A multivariate statistical assessment using profile analysis. Fifth Scientific Meeting of the International Society for Magnetic Resonance in Medicine, Vancouver, B.C.: ISMRM. 354, 1997.

R. L. Somorjai, M. Jarmasz, and R. Baumgartner. A Fast, Two-Stage Strategy for Exploratory Analysis of Functional MRI Data by Temporal Fuzzy Clustering. CIMA2001, Bangor, Wales, 2001.

R. L. Somorjai, M. Jarmasz, R. Baumgartner, and W. Richter. Exploratory analysis of fMR images: Voxel preselection via "self-similarity." Proceedings of HBM, Düsseldorf, 46, 1999.

R.L. Somorjai. Meta Brain, Meta Experiment - Simultaneous Analysis of fMRI Data Across Multiple Subjects and / or Experiments. 7th Int. Conf. on Functional Mapping of the Human Brain, Brighton, NeuroImage 13, 6, S253, 2001.

Sowa M.G., Mansfield J.R., Scarth G.B., Mantsch H.H. (1997) Non-invasive assessment of regional and temporal variations in tissue oxygenation by near infrared spectroscopy and imaging. Appl Spectr 51, 143-152.

A. Taleb and C. Jutten. Nonlinear source separation: the post-nonlinear mixtures. Proc. 1997 European Symp. on Artificial Neural Networks, Bruges, Belgium, 279-284. 1997.

G. Tian, J. Shen, J. Sun, B. Xiang, G.I. Oriaku, L. Zhezong, G. Scarth, R.L. Somorjai, J.K. Saunders, T.A. Salerno, and R. Deslauriers. Does simultaneous antegrade/retrograde cardioplegia improve myocardial perfusion in the areas at risk? A magnetic resonance perfusion imaging study in isolated pig hearts. J. Thorac. Cardiovasc. Surg. 115, 913-924, 1998.

J. T. Tou and R. C. Gonzales. Pattern Recognition Principles. Addison-Wesley Publ. Co., Reading, MA, 1974.

J. W. Tukey. The future of data analysis. Annals of Statistics 33, 1-67, 1962.

J. W. Tukey. Exploratory Data Analysis. Addison-Wesley, Reading, MA, 1977.

U. I. Tuor, K. Malisza, T. Foniok, R. Papadimitropoulos, M. Jarmasz, R. L. Somorjai, and P. Kozlowski. Functional Magnetic Resonance Imaging in Rats Subjected to Noxious Electrical and Chemical Stimulation of the Forepaw. Pain 87, 315-324, 2000.

E. J. Wegman. Huge data sets and the frontiers of computational feasibility. J. Computational and Graphical Statistics 4, 281-295, 1995.

E.J. Wegman. On the eve of the 21st century: statistical science at a crossroads. Computational Statistics and Data Analysis 32, 239-243, 2000.

E. J. Wegman. Visions: New techniques and technologies in statistics. Computational Statistics 15, 133-144, 2000.

H. Yang, S. Amari, and A. Cichocki. Information-theoretic approach to blind separation of sources in non-linear mixture. Signal Processing 64, 291-300, 1998.

J. Ye, L. Yang, M. R. Del Bigio, R. Summers, D. Jackson, R. L. Somorjai, T. A. Salerno, and R. Deslauriers. Retrograde cerebral perfusion provides limited distribution of blood to the brain: a study in pigs. J.Thorac.Cardiovasc.Surg. 114, 660-665, 1997.

Additional information on EvIdent®, with FAQs, is available on www.ibd.nrc.ca/informatics

3 Testing Competing Hypotheses about Single Trial fMRI

Lars Kai Hansen
Department of Mathematical Modelling
Technical University of Denmark B321
DK-2800 Lyngby, Denmark
lkhansen@imm.dtu.dk

Archana Purushotham and Seong-Gi Kim
Center for Magnetic Resonance Research,
Dept. of Radiology and Dept. of Biomedical Engineering,
University of Minnesota, USA.

We use a Bayesian framework to compute probabilities of competing hypotheses about functional activation based on single trial fMRI measurements. Within the framework we obtain a complete probabilistic picture of competing hypotheses, hence control of both type I and type II errors.

3.1 Introduction

In single trial fMRI experiments we may want to analyze the local activation with respect to several competing hypotheses. As a specific example we have designed an experiment which consists of a cued, delayed motor action with a delay between cue and go-signal that varies from trial to trial. In each pixel we then face three competing hypotheses: *no activation*, *motor preparation* (the activation last for the delay between cue and action), and *motor execution* (the activation lasts only for the time interval of the actual execution). In figure 3.1 we show the two activation reference functions corresponding to *motor preparation* and *motor execution* hypotheses. The reference functions are binary time series, the actual fMRI signal will be modeled below as the reference function convolved with a linear filter (the hemodynamic response) degraded by additive white noise.

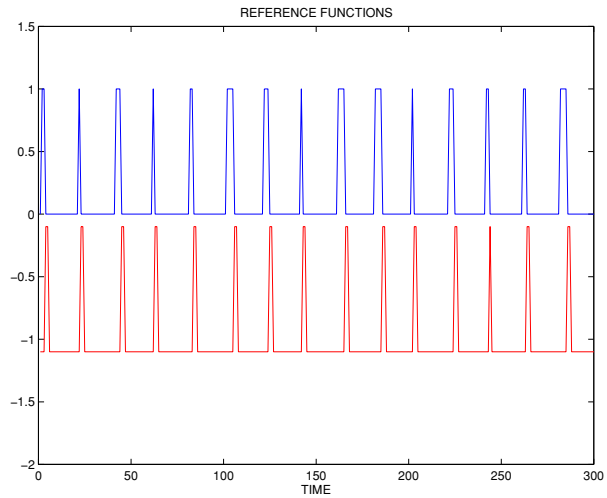


Figure 3.1 The two reference functions for a sequence of single trial cued, delayed motor activation experiments. The two reference functions have been offset vertically for illustration, the upper function is "on" during the preparation phases, while the lower reference function is on during the execution phases.

We will present a Bayesian framework below that allows calculation of relative probabilities of such competing hypotheses. The Bayesian framework is of interest in this context because it gives a more complete picture of the interplay between null hypotheses and alternatives and the framework has embedded a quantitative statement of the a priori knowledge that enters the formulation of hypotheses.

Frank et al. (1998) recently reviewed a Bayesian framework for signal detection in fMRI data. Here we expand on the application of the Bayesian framework based on so-called *conjugate priors* and we include an explicit treatment of linear hemodynamic response effects.

3.2 Bayes's Theory

We will be concerned with models of the local activation in a region or a single pixel. Let $\mathbf{y}$ be a fMRI signal measured at times $t = 1, ..., T$, and represented as a $T \times 1$ vector with components $\mathbf{y}(t)$. The experiment is characterized in terms of one or more activation reference functions, representing alternative hypotheses about the local activation. Consider linear fMRI signal models of the form,

$$\widehat{\mathbf{y}}(t) = \sum_{\tau=0}^{l} \mathbf{x}(t-\tau)\mathbf{b}(\tau) \tag{3.1}$$

where $\mathbf{x}(t)$ is the average activation in the region or pixel under consideration at time t. $\mathbf{b}$ is a set of $(l+1)$ linear coefficients describing the local hemodynamic response to the activation. Introducing the $T \times (l+1)$ matrix with components

$\mathbf{X}(t,\tau) \equiv \mathbf{x}(t-\tau)$ the linear model can be written in matrix form

$$\widehat{\mathbf{y}} = \mathbf{X}\mathbf{b}. \tag{3.2}$$

In an fMRI experiment we expect that the actual measurement deviates from the "ideal" model output by various noise contributions that we will represent by a random white noise process so that $\mathbf{y}(t) = \widehat{\mathbf{y}}(t) + \mathbf{n}(t)$ where $\mathbf{n}(t)$ is assumed zero mean normal with unknown variance (σ^2). Such models have been investigated e.g. by Lange and Zeger (1997) based on parameterized hemodynamic response functions $\mathbf{b}(\tau)$ of the gamma density form. In (et al., 1997) more general finite impulse response models (FIR models) were invoked. FIR models fit the coefficients $\mathbf{b}(\tau)$ in the hemodynamic response function individually. A fundamental problem with equation (3.2) is that the neither the model "order" l nor the noise variance σ^2 are known.

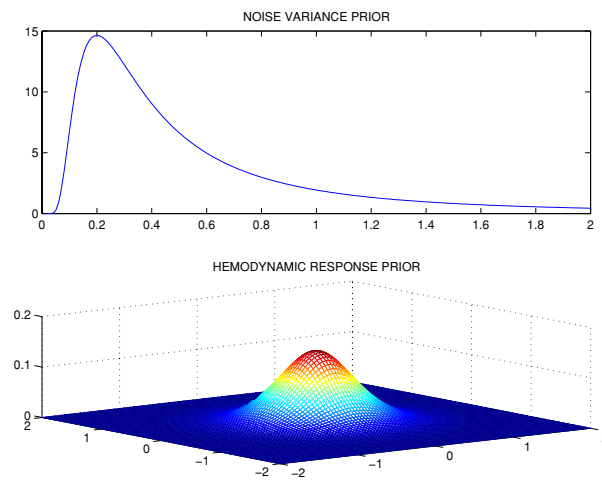


Figure 3.2 Visualization of the priors on the noise variance and the coefficients of the hemodynamic impulse response. The upper panel shows the inverse gamma distribution of the noise variance parameter, while the lower panel show the joint distribution of two hemodynamic impulse response parameters. The parameters of both distributions are matched to the actual paradigm and expected signal-to-noise levels.

A main objective of neuroimaging is to answer questions about the form of activation in regions or locations in the brain. When testing competing hypotheses about the activation of a certain region (or a single pixel) we consider several possible activation patterns $\mathbf{X}_m$ indexed by m. We will show below how we can compute the probability for a given model order, and the probability of a given type of activation (m) given the measurement $\mathbf{y}$ is then

$$P(m|\mathbf{y}) = \sum_{l=1}^{l_{\max}} P(m,l|\mathbf{y}) = \sum_{l=1}^{l_{\max}} P(\mathbf{y}|m,l)P(m,l)/P(\mathbf{y}), \tag{3.3}$$

where we have used Bayes' relation and where $P(m,l)$ is a *prior* probability of the combination of reference function and model dimension. In the following we will let the prior probabilities be a given by a uniform distribution over combinations of (m,l), in the set of reference functions and the corresponding model orders $(l=1,...,l_{\max})$, hence, we express no *a priori* preference for any of the alternatives.

For a fixed set of parameters $\mathbf{b},\sigma^2$ we can use equation (3.2) to establish the likelihood function, i.e., the probability density of the observation s given the parameters,

$$P(\mathbf{y}|\sigma^2,\mathbf{b},\mathbf{X},l)=\left(\frac{1}{2\pi\sigma^2}\right)^{T/2}\exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\mathbf{b})^2\right). \tag{3.4}$$

Since, however, these parameters too are unknown we need to eliminate them using a *prior distribution* $P(\mathbf{b},\sigma^2|m,l)$ which quantifies the general knowledge we have on the domain and which potentially depends on the given activation function and model order. Following Ohagan (1994) one can write

$$\begin{aligned}P(\mathbf{y}|m,l)&=\int d\sigma^2\int d\mathbf{b}P(\mathbf{b},\sigma^2|m,l)P(\mathbf{y}|\sigma^2,\mathbf{b},\mathbf{X}_m,l)\\&=\int d\sigma^2\int d\mathbf{b}P(\mathbf{b},\sigma^2|m,l)\left(\frac{1}{2\pi\sigma^2}\right)^{T/2}\exp\left(-\frac{(\mathbf{y}-\mathbf{X_m}\mathbf{b})^2}{2\sigma^2}\right).\end{aligned} \tag{3.5}$$

We will use the principle of *congugate priors* to establish a convenient prior $P(\mathbf{b},\sigma^2|m,l)$ (Ohagan, 1994). The conjugate prior for the above linear model with additive gaussian noise is the so-called *normal-inverse-gamma* or $\mathrm{NIG}(a,d,\mathbf{m},\mathbf{V})$, distribution,

$$P(\mathbf{b},\sigma^2|m,l)=\frac{(a/2)^{d/2}(\sigma^2)^{-(d+l+2)/2}}{(2\pi)^{l/2}|\mathbf{V}|^{1/2}\Gamma(d/2)}\exp\left(-(\mathbf{b}-\mathbf{m})'(2\sigma^2\mathbf{V})^{-1}(\mathbf{b}-\mathbf{m})-\frac{a}{2\sigma^2}\right). \tag{3.6}$$

The new (hyper-) parameters $d,a,\mathbf{m},\mathbf{V}$ have the following meaning. The marginal prior distribution of $\mathbf{b}$,

$$\begin{aligned}P(\mathbf{b}|m,l)&=\int d\sigma^2 P((\mathbf{b},\sigma^2|m,l)\\&=\frac{(a/2)^{-l/2}\Gamma((d+l)/2)}{(2\pi)^{l/2}|\mathbf{V}|^{1/2}\Gamma(d/2)}\left(1+(\mathbf{b}-\mathbf{m})'(a\mathbf{V})^{-1}(\mathbf{b}-\mathbf{m})\right)^{-(d+l+2)/2}\end{aligned} \tag{3.7}$$

is a multivariate t-distribution with mean $\mathbf{m}$ and the covariance determined by $(a/(d-2))\,\mathbf{V}$. This distribution is unimodally centered at $\mathbf{m}$, with heavier "tails" than a normal distribution, see figure 3.2. The marginal prior distribution of σ^2 is given by

$$P(\sigma^2|m,l)=\frac{(a/2)^{-d/2}(\sigma^2)^{-(d+2)/2}}{\Gamma(d/2)}\exp\left(-a/(2\sigma^2)\right). \tag{3.8}$$

Hence an inverse gamma distribution (meaning: $1/\sigma^2$ is gamma distributed) of mean $a/(d-2),\ d>2$.

The next step of the inference is then to set the parameters of the prior. In general we prefer to give the parameters values so that they have minimal influence on results. In particular, we should check that for long time series their effects should vanish completely. The prior mean of the noise variance can, e.g., be set to the observed signal variance, $a/(d-2) = \sigma_y^2 \equiv \mathbf{y}'\mathbf{y}/T$, meaning that we do not expect much larger noise variance than the total observed variance. Further we will let $d = 3$ leading to a prior as shown in figure 3.2. We express no prior knowledge about the mean hemodynamic response, hence, $\mathbf{m} = \mathbf{0}$. The form of the prior covariance structure will be $\mathbf{V} = v\mathbf{1}$, where $\mathbf{1}$ is a unit matrix. The parameter v will be determined essentially by data. The prior variance of the fitted signal $\widehat{\mathbf{y}}$, is given by

$$\langle \widehat{\mathbf{y}}'\widehat{\mathbf{y}} \rangle_{\text{prior}}/T = \mathsf{Tr}[\mathbf{X}\mathbf{X}'\langle \mathbf{b}\mathbf{b}' \rangle_{\text{prior}}/T] = (va/(d-2))\mathsf{Tr}[\mathbf{X}\mathbf{X}']/T \tag{3.9}$$

We can let this be some fraction Q of the variance of the measured signal, i.e., let $v = Q/\mathsf{Tr}[\mathbf{X}\mathbf{X}']/T \sim Q/((l+1)\sigma_{x,m}^2)$, where $\sigma_{x,m}^2$ is the variance of the m'th reference function. The role of the parameter Q will be investigated below; we find that there is a rather wide window of values of this parameter in which the Bayes decisions are stable.

Comparing equations (3.2) and (3.6) we see that by conjugacy they are of the same exponential form, so when we multiply them together the function to be integrated in equation (3.5) is again an (un-normalized) NIG distribution, hence the integral is simply the NIG normalization integral (Ohagan, 1994), we find

$$P(\mathbf{y}|m,l) = \left(\frac{|\mathbf{V}_{\mathrm{P}}| a^d}{|\mathbf{V}| (a_P)^{d_P} \pi^T} \right)^{1/2} \frac{\Gamma(d_P/2)}{\Gamma(d/2)}, \tag{3.10}$$

with the following definitions

$$\mathbf{V}_P^{-1} = \mathbf{V}^{-1} + \mathbf{X}'\mathbf{X}, \tag{3.11}$$

$$\mathbf{m}_P = \mathbf{V}_P(\mathbf{V}^{-1}\mathbf{m} + \mathbf{X}'\mathbf{y}), \tag{3.12}$$

$$a_P = a + \mathbf{m}'\mathbf{V}^{-1}\mathbf{m} + \mathbf{y}'\mathbf{y} - \mathbf{m}'_P\mathbf{V}_P^{-1}\mathbf{m}_P, \tag{3.13}$$

$$d_P = d + T. \tag{3.14}$$

Using our specifications of the prior parameters we obtain the simplification

$$\mathbf{V}_P^{-1} = \frac{(l+1)\sigma_{x,m}^2}{Q}\mathbf{1} + \mathbf{X}'\mathbf{X}, \tag{3.15}$$

$$a_P = (T+1)\sigma_y^2 - \mathbf{y}'\mathbf{X}\mathbf{V}_P\mathbf{X}'\mathbf{y}, \tag{3.16}$$

$$d_P = 3 + T. \tag{3.17}$$

We can see explicitly that the influence of the prior is weak for $T \gg 1$.

Testing the above linear system hypotheses, a natural null-hypothesis is formed by $\mathbf{x}_0 \equiv 0$, corresponding to a *no fitted signal* model. The corresponding probability density is given by the $\mathbf{X} = \mathbf{0}$ limit of the above expressions.

Assuming that all models are equally probable *a priori*, we get the appropriately normalized posterior probabilities,

$$P(m|\mathbf{y}) = \frac{\sum_l P(\mathbf{y}|m,l)}{P(\mathbf{y}|0) + \sum_{m,l} P(\mathbf{y}|m,l)}$$
$$P(0|\mathbf{y}) = \frac{P(\mathbf{y}|0)}{P(\mathbf{y}|0) + \sum_{m,l} P(\mathbf{y}|m,l)} \tag{3.18}$$

In our particular cued, delayed motor action experiment there will be three such probabilities.

3.3 Evaluation on Simulated and fMRI Data

3.3.1 Simulation Experiments

In order to illustrate the viability of the Bayesian approach for testing hypotheses about time series data we have set up simulation experiments. The experiment involves a stimulus reference function, here taken to be a simple on/off block design for simplicity. The temporal extent of the experiment is $T = 200$ time units (TR's). Figure 3.3 shows the stimulus reference function, the hemodynamic response obtained by convolving the impulse response (here a *boxcar* of duration $l_0 = 10$ TR's, see figure 3.5) and the reference function, and finally we show the noise degraded (simulated) fMRI signal.

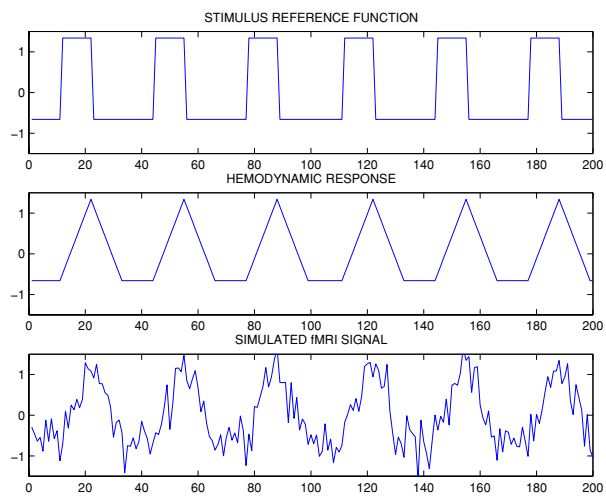


Figure 3.3 Simulation experiment A. The upper panel shows the stimulus reference function for a simple block design. The hemodynamic response function is here modeled as a boxcar of length 10 (TR's). The resulting hemodynamic response is shown in the second panel, and finally we show the simulated fMRI signal in the lower panel. The noise level is 0.3 (ratio of standard deviations – noise/signal).

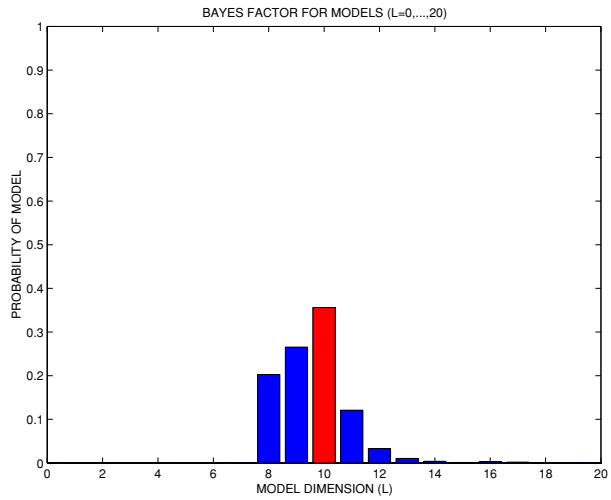


Figure 3.4 Simulation experiment A. The stimulus reference function for a simple block design as shown in figure 1. The hemodynamic response function is here modeled as a boxcar of length 10 (TR's) see figure 3. The model probability was computed for models with lags ranging from 0 to $l_{\max}, l_{\max} = 20$. The probability of the zero lag model (no fitted signal) is the nul hypothesis. The figure shows the probability of each of the 21 models considered and is as expected focused on the "correct" model $l = 10$

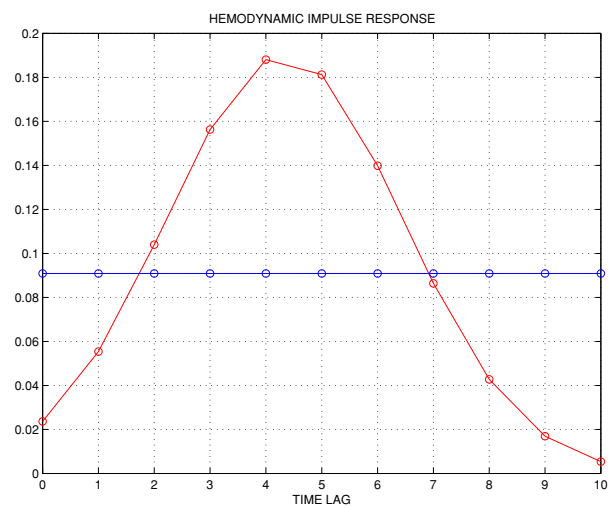


Figure 3.5 Two model hemodynamic response functions used in the initial experiments. The hemodynamic response function is here modeled either as a boxcar or a Gaussian of both length 10 (TR's).

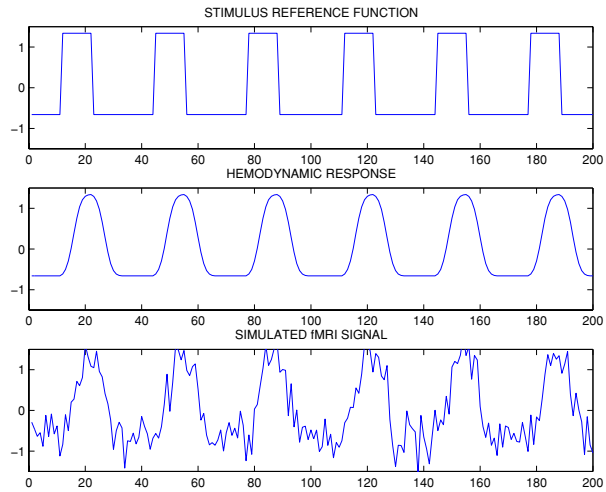


Figure 3.6 Simulation experiment A. The upper panel shows the stimulus reference function for a simple block design. The hemodynamic response function is here modeled as a Gaussian of length 10 (TR's). The resulting hemodynamic response is shown in the second panel, and finally we show the simulated fMRI signal in the lower panel. The noise level is 0.3 (ratio of standard deviations – noise/signal).

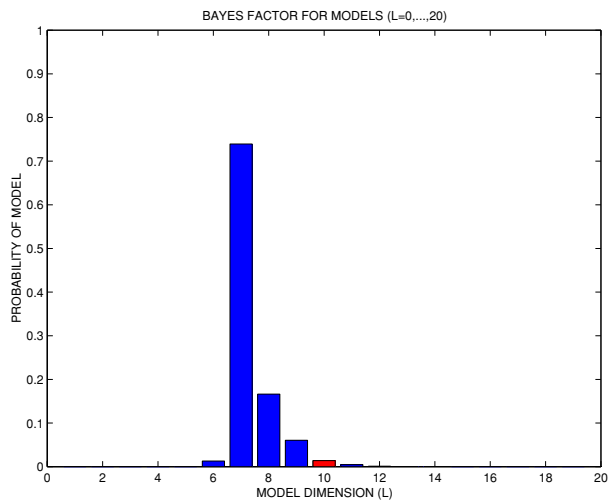


Figure 3.7 Simulation experiment A. The stimulus reference function for a simple block design as shown in figure 1. The hemodynamic response function is here modeled as a boxcar of length 10 (TR's). The probability $P(\text{model}|\text{data})$ was computed for models with lags ranging from 0 to $l_{\max}, l_{\max} = 20$. The probability of the zero lag model is the null-hypothesis probability (no fitted signal). The figure shows the probability of each of the 21 models considered, since the Gaussian model has small (relative to the noise level) contributions to the fitted signal for large delays , the probability is focused on somewhat smaller models than the "correct" model $l = 10$

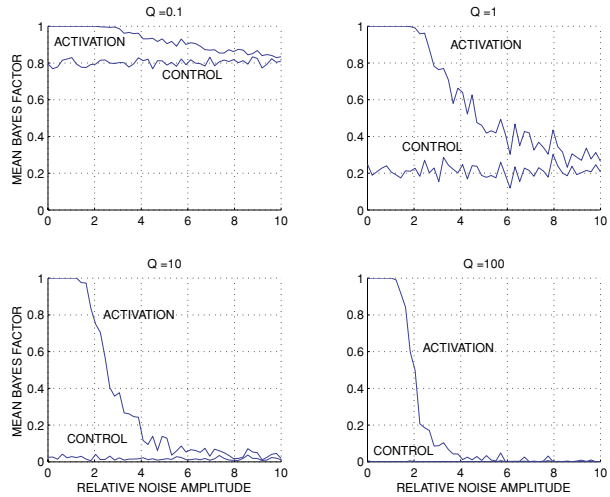


Figure 3.8 Simulation experiment B. The stimulus reference function for a simple block design as shown in figure 1. The hemodynamic response function is here modeled as a Gaussian of length 10 (TR's). The model probabilities $P(\text{model}|\text{data})$ were computed for models with lags ranging from 0 to $l_{\max}, l_{\max} = 20$. The zero lag model is the $\mathbf{X} = \mathbf{0}$ null-hypothesis (the measured signal is modeled as a white noise Gaussian signal with unknown variance). We calculated the probability of the activation (summing again over all model "lags"), for increasing noise levels. In the plot we show the probability for four different Q's and in each case we show the probability as function of the relative noise amplitude for a situation where the signal actually was activated (*ACTIVATION*) and for the case when the reference function is zero (*CONTROL*). We note that for Q's in the range $10 - 100$ we find good detection of signal for relative noise levels less than 1 (low type-II error rate), and at the same time good suppression of the activation probability in case of no activation (*CONTROL*), i.e., low type-1 error rate.

In simulation experiment A we test the ability of the Bayesian approach to find the correct model, i.e., the correct duration of the response. In figure 3.4 we show the Bayes probabilities $P(l|\text{data})$, for $l = 0, ..., l_{\max}$, with $l_{\max} = 20$. As seen the probabilities are indeed centered on the correct value $l \sim 10$.

The boxcar filter is, of course, not a very realistic model of the hemodynamic response, so the above experiment was redone in figures 3.6 and 3.7 for the *Gaussian* impulse response shown figure 3.5. Since the Gaussian shape of the impulse response has only little weight for the larger delays, the probabilities are centered on somewhat smaller values of l.

Simulation experiment B investigates the influence of the prior expected signal-to-noise level, that is, the role of the parameter Q. Figure 3.8 displays the probability of activation (summed on all filter lenghts) for both data with activation (*ACTIVATION*) and without activation (*CONTROL*) for four different Q-values. It is seen that for Q's in the range $10 - 100$ we get good detection for reasonable relative noise levels, hence, low type-II error rate, while at the same time suppression of the activation probability for signals without activation, hence low type-I error rate.

3.3.2 fMRI Experiment

We now return to the experiment briefly introduced in the introduction. We analyze a small subset of an experiment designed to determine the location of two different aspects of motor control (see (Purushotham et al., 2001) for details and analysis of the complete data set). The experiment involves a cued, delayed, motor response and in the motor areas we can expect different behavior in regions involved in planning versus execution of the response. In the context of the new method we consider here three competing hypotheses. A pixel can either be non-activated (null-hypothesis), be activated during preparation or be activated during execution.

The paradigm consisted of a delayed, cued joystick movement task performed by normal, right-handed adult human subjects. The subject was asked to move a cursor from the centre of the screen to the memorized location of a target using a joystick. The cursor was positioned in the centre of the screen at the beginning of the task. A yellow target appeared for 200 ms at one of 8 possible locations along the circumference of a circle centered on the screen (view angle: < 5 degrees). Then the target disappeared. Following a pseudo-randomized variable (0-3 seconds) delay period, the colour of the central circular home zone changed from red to green, the *go* cue. The subject was to then move the cursor radially to the memorized location of the target, as accurately and quickly as possible, and then move it back to the home circle. This signaled the end of a single trial, and a return to the control state. During the control period between tasks, the targets continued to appear and disappear, one every 3 seconds, to control for visual stimulation and eye movement. The home circle remained black during this period, and the subject was instructed not to make any joystick movements.

A modified joystick (Measurement Systems, Model 521) from which all ferromagnetic components were removed, was used. The maximum range of motion is an arc of $+/-30$ degrees. The joystick was positioned and secured to the side of the subjects right thigh, at a position that was comfortable for the subject to reach with the right hand while lying down. Subjects moved the joystick using their right thumbs and forefingers in a pincer grasp. The output of the joystick was sampled at 100 Hz, and controlled the movement of the cursor on the screen.

16 single trials as described above, were performed, with an interval of 20 sec between the beginning of one trial and the next 20 volumes (at a rate of 1 volume per sec) were acquired per trial, and along with 20 initial pre-steady-state volumes, made up a total of 340 volumes per run. Response accuracy, reaction and movement times and trajectories were recorded for each trial but were not used in the present analysis.

3.3.2.1 Data Acquisition

Magnetic resonance imaging experiments were performed on a 4 Tesla whole body MRI system (Varian, Palo Alto, CA) with a homogeneous birdcage coil. Multi-slice T1-weighted scout images were acquired with a Turbo FLASH sequence (inversion

time of 1.2 s) to select the slices for functional imaging. From these anatomical images, ten coronal 5mm slices, including the primary, pre- and supplementary motor areas, were selected for the joystick task. For the mental rotation task, four axial 10 mm slices were chosen, including the primary and secondary motor areas. For functional MRI studies, a gradient-echo echo-planar imaging (EPI) technique was used. Typical fMRI parameters were matrix size of 64 x 64, field of view (FOV) of 24 x 24 cm2, echo time (TE) of 20 ms (for joystick experiment) or 15ms (for mental rotation experiment), and repetition time of 1000 (for joystick experiment) or 480 ms (for mental rotation experiment).

The study necessitated the provision of visual instruction, and stimuli were generated by a PC. The PC was synchronized with the MRI data collection. Images projected onto a screen inside the magnet room via a projector were viewed by subjects through a mirror. The screen subtended less than 5 degrees of the field of view.

3.3.2.2 Testing Competing Hypotheses

In each voxel we test three hypotheses: a null, and two different reference activation functions, demarcating the preparation and the execution phases of the response respectively, as shown in figure 3.1.

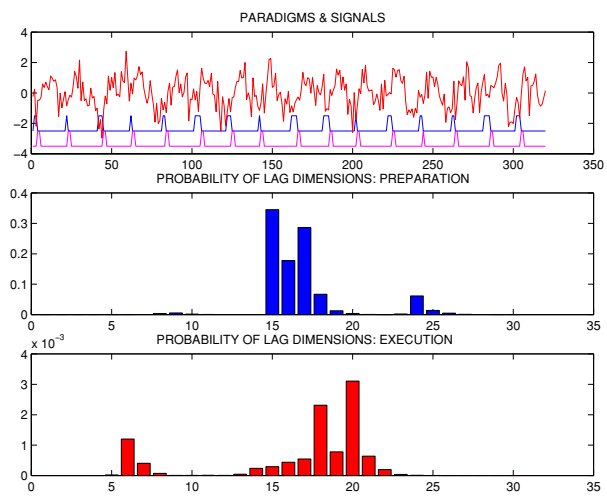


Figure 3.9 Distribution of probabilities across different model lags ($l = 1, .., 30$) for two models based on different reference functions in the region of a single pixel. In the upper panel we show the measured fMRI signal and the two reference functions for a cued-delay motor activation single trial experiment (from above: fMRI signal, motor preparation, and motor execution). Hypothesis $m = 1$ is a reference function corresponding to motor preparation, while $m = 2$ corresponds to motor execution. This pixel shows high probability of the preparation model.

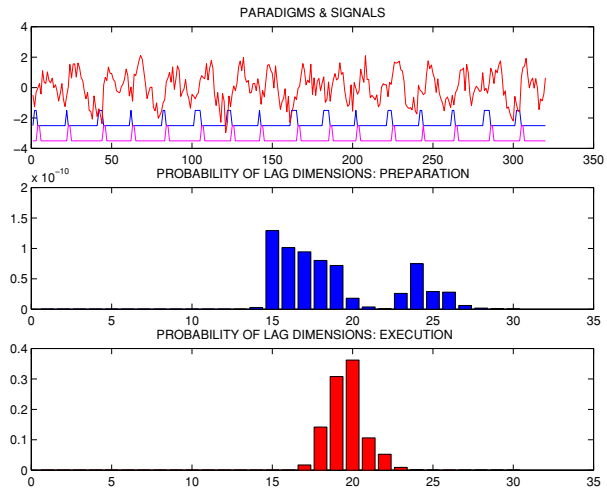


Figure 3.10 Distribution of probabilities across different model lags ($l = 1, .., 30$) for two models based on different reference functions. In the upper panel we show the measured fMRI signal, and the two reference functions for a cued-delay motor activation single trial experiment (from above: fMRI signal, motor preparation, and motor execution). Hypothesis $m = 1$ is a reference function corresponding to motor preparation, while $m = 2$ corresponds to motor execution. This pixel shows high probability of the execution model.

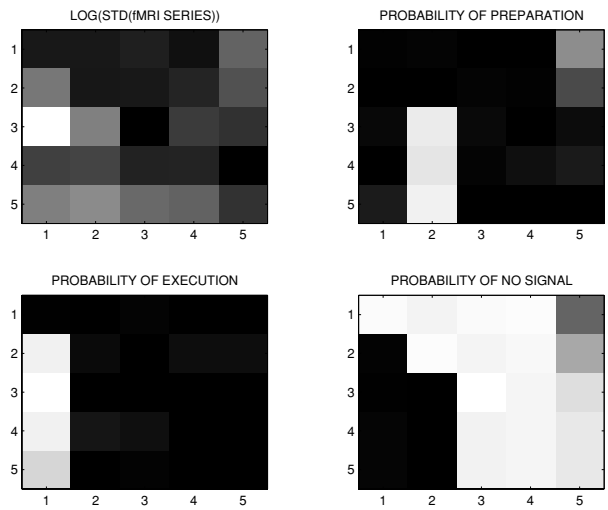


Figure 3.11 Distribution of probabilities in the 5×5 slap of pixels covering pre-motor and motor areas. In the upper left corner we show the signal standard deviation σ_y for the 25 pixels. The three other panels shows the spatial distribution of the probability for the two activation hypotheses and for the null-hypothesis. We use a linear gray map so that bright pixels have probability ~ 1, dark pixels have probability ~ 0).

For each voxel in a 5-by-5 slab of pixels expected to cover motor and pre-motor areas aggregate probabilities (averaged over both response values and response lengths) for the where calculated and presented in figure 3.11.

The probabilities suggest that all three types of response (null, preparation and execution) are present in the slab. For illustration of the Bayesian analysis we also present the probability distributions over response lengths in figures 3.9-3.10 for both the preparation hypothesis and the execution hypothesis. The two pixels have strongly peaked probability distributions for the two different hypotheses respectively at response durations of 15 and 20 TR's.

3.4 Conclusion

We have outlined a Bayesian framework for signal detection in noisy linear systems. We used weak conjugate priors and as a result we obtain closed form expressions for the relative probabilities over competing hypotheses, depending on only one free parameter. The value of this parameter can be estimated from simulations and the system appears to quite insensitive to its precise value. We used the Bayesian framework to estimate the probability of three alternatives in an fMRI experiment involving planning and execution of motion. In a small slap covering both motor and premotor pixels we found contiguous regions designated to the null, the execution and to the preparation hypotheses.

Acknowledgments

This work was funded by the Danish Research Councils through the THOR Center for Neuroinformatics, and by the National Institutes of Health, USA, grants MH57180 and RR08079.

References

F. Å. Nielsen, L. K. Hansen, P. Toft, N. Goutte, S. C. Strother, N. Morch, C. Svarer, R. Savoy, B. Rosen, E. Rostrup, and P. Born. Comparison of two convolution models for fmri time series. *NeuroImage*, 5:473, 1997.

L. R. Frank, R. B. Buxton, and E. C. Wong. Probabilistic analysis of functional magnetic resonance imaging data. *Magnetic Resonance in Medicine*, 39:132–148, 1998.

N. Lange and S. L. Zeger. Non-linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging. *J. Royal Stat. Soc, Appl. Stat.*, 46:1– 26, 1997.

A. Ohagan. *Bayesian Inference*. The University Press, Cambridge, 1994.

A. Purushotham, F. Å. Nielsen, L. K. Hansen, and S-G. Kim. Resolving activity due to different functional components of a cognitive task in fmri single trial data using meta-k-means clustering. *Unpublished*, 2001.

4 Deterministic and Stochastic Features of fMRI Data: Implications for Data Averaging

Martin J. McKeown
Depts of Medicine (Neurology) & Biomedical Engineering
Brain Imaging and Analysis Center (BIAC)
Center for Cognitive Neuroscience
Box 3918
Duke University Medical Center
Durham, NC, USA
martin.mckeown@duke.edu

Averaging of data, using time windows time-locked to repetitive stimuli, is a method that has long been used in Event-Related Potential (ERP) research. As event-related designs are becoming increasingly common in fMRI experiments, selective averaging is a natural approach to the analysis of these data sets. However, as the biophysical origin (and presumably the statistical properties) of the fMRI BOLD and ERP signals fundamentally differ, there is a need to assess the implications of averaging raw fMRI data. We recorded a fMRI data series from a single subject performing a simple event-related task, consisting of 95 presentations of checkerboard visual stimuli. The data set was first dimension-reduced with Principal Component Analysis (PCA) and separated into 100 spatially independent components with Independent Component Analysis (ICA), an iterative technique whose weight matrix is normally initialized to the identity matrix. To determine components which were reproducible, and by inference represented deterministic features in the data, the ICA processing step was repeated, but this time initialized with the inverse of the weight matrix computed from the first analysis, a method supported by simulations. The mutual information between best-matching pairs of components each ICA analysis was plotted. Visual inspection suggested that 55 components were reproducible, accounting for 84% of the variance in the dimension-reduced data. The reproducible components exhibited much less trial-to-trial variability than the raw data from even the most activated voxels. Of the 55 reproducible components

in the first series, the average responses of 28 independent components were significantly affected by stimulus presentation ($p < 0.001$). The most significant stimulus correlated component was strongly time-locked to stimulus presentation and was directly stimulus correlated, corresponding to occipital brain regions. Other stimulus correlated components included stimulus-correlated motion artifact, overlapping occipital activation with a different time course from the strongly time-locked component, and frontal and temporal activations. Our results suggest that a significant proportion of the variance in fMRI data is in fact deterministic. Double ICA training with different initial weight matrices is a simple and practical method to determine which components are reproducible. Averaging the time courses of robust spatially-independent components time-locked to stimulus presentation, as opposed to the raw data, may prevent possible biases in the estimates of the spatial and temporal extent of stimulus correlated activation and of trial-to-trial variability.

4.1 Introduction

Functional Magnetic Resonance Imaging (fMRI) experiments usually incorporate "block" designs or "event-related" designs. With block designs, stimulus events are clumped together within an extended 15–40 sec block, interspersed with blocks of a contrast condition, such as rest in a motor study. Several interspersed blocks are typically performed, to prevent slow drifts in the fMRI signal from being misinterpreted as true changes due to brain activation. With event-related designs, e.g. (Buckner et al., 1998; McCarthy et al., 1996), stimuli are presented singly and sufficiently separated in time so that the mean latency, rise, amplitude, duration, and recovery in response to the experimental stimulus can be estimated.

A major disadvantage of event-related designs is that the Blood Oxygen Level Dependent (BOLD) response to an impulse function of neural activity is long, necessitating interstimulus intervals of the order of 15s to avoid overlapping responses from successive events (Bandettini and Cox, 2000; Howseman et al., 1998). Usually post-processing methods, based on linear, time-invariant models of evoked response (Boynton et al., 1996; Buckner et al., 1998) are used to estimate the mean effect of a single stimulus (Buckner et al., 1996, 1998; Rosen et al., 1998; Burock et al., 1998). However, there is evidence for non-linearities under realistic trial conditions, e.g., (Vazquez and Noll, 1998; Huettel et al., 2001).

Analysis of data from fMRI experiments incorporating event-related designs typically involves dividing the data from each voxel into epochs, time-locked to stimulation presentation, which are then averaged to obtain a mean response to the stimuli. This method of analysis, standard for analysis of Event Related Potential (ERP) data (McCarthy, 1999), implicitly assumes that the data can be accurately modeled as a deterministic signal precisely time-locked to stimulus presentation, and corrupted with random noise that will tend to zero when averaged over many trials. All brain signals not precisely time-locked to stimulus presentation are handled as noise with this model. However, if underlying components of the data are not

completely random with respect to the onset of stimulus presentation, they may tend to average to values other than zero, introducing biases in the estimates of stimulus-locked signals.

Unlike ERP studies, the maximum number of non-overlapping trials that can be employed in an fMRI study is practically limited to a few hundred across all experimental conditions, and is usually much less. Even if a feature of the fMRI data would eventually tend to zero when averaged over hundreds of trials, this may not necessarily be the case when averaged over the more common case of 30–40 trials per condition. A further problem is that, in contrast to EEG data, characterization of artifacts such as cardiac and respiratory pulsations and movement are less well known and not as readily identified.

Accurate isolation of non-stimulus correlated but non-random aspects of fMRI data is a worthwhile goal, as adequately modeling their effects is a first step to minimizing their influence on estimates of stimulus correlated brain activity. For example, in a standard regression approach using the General Linear Model, (Friston, 1996) the total variance of the data is assumed to be appropriately allocated into deterministic and random components. The deterministic fraction may contain, as nuisance regressors, non-brain activity such as cardiac pulsations. Under the limitations imposed by the statistical assumptions of the model, the contribution of these nuisance regressors can be stripped away from the raw data to more accurately reveal underlying hypothesized stimulus correlated activity.

Some clues suggest that fMRI data is not heavily corrupted with Gaussian noise, although this is a standard feature of many models. McKeown and Sejnowski (1998), by calculating the log-likelihood of observing the data under an assumed linear model (Independent Component Analysis, ICA), demonstrated that true fMRI data is unlike what would be expected by a relatively small number of deterministic components corrupted by zero-mean Gaussian distributed random noise.

In a related study, training on alternate even and odd time points of an fMRI data or corrupting the data with pure Gaussian noise did not significantly affect the estimates of stimulus correlated component of interest, suggesting the stimulus correlated component was deterministic and reproducible (McKeown et al., 1998).

There are many reasons why fMRI data may exhibit variability to a given stimulus, including non-reproducibility of the brain's response to stimulus, ongoing brain activity not related to stimulus, differential hemodynamic responses to the same neuronal activity, and motion artifacts. Here we differentiate between deterministic and random variability. In this chapter, we refer to variability as deterministic if the state of a dynamical system at selected time points or points in space can be modeled predictably from knowledge of the system at other time points or other points in space. In contrast, we define a signal as random if knowledge of the signal at some time points or points in space provide no information (i.e. is statistically independent from) the signal at other points in time or space.

Here we exploit two facts regarding the Bell-Sejnowski ICA algorithm: 1) when the assumptions of are perfectly held, the separated ICs are unique and 2) that the algorithm cannot separate pure Gaussian noise. We propose that the reproducibility

of the separated ICs will be a qualitative indication of the amount the data are corrupted with truly random noise.

By calculating independent components with different initial weight matrices, we will determine which components are robust to changes in weight matrix initialization. We will infer that these components represent deterministic features of the data. We show that the contribution of some robust components, presumably based on endogenous brain signals, do not sum to zero when averaged time-locked to stimulus presentation after a reasonable number of trials, and have the potential to introduce biases into the temporal and spatial estimates of stimulus-locked activity.

When separating fMRI data into spatially-independent components, the ICA model is,

$$C = WX \tag{4.1}$$

where C is an n x v matrix of component maps (where n is the number of time points in the experiment and v is the number of brain voxels), X is an n x v row mean-zero data matrix, with each row representing the entire volume recorded at each given time point, and W is an n x n unmixing matrix, usually iteratively obtained. Since the number of time points in a given experiment (especially event-related designs) may greatly exceed the effective dimensionality of the data, it is common to first reduce the data to a linear subspace with principal component analysis (PCA) before applying the ICA algorithm (McKeown et al., 1998):

$$X_{red} = V^T X \tag{4.2}$$

where V is n by p ($p < n$) matrix whose columns are the eigenvectors of the covariance matrix, $< XX^T >$, corresponding to the p largest eigenvalues, and V^T is transpose of V. X_{red} is a now smaller full-rank matrix of *eigenimages* of X. ICA decomposition of the resulting eigenimages, X_{red}, gives,

$$C_{red} = WX_{red} \tag{4.3}$$

where C_{red} is the p by v matrix of component maps, and W is the p by p computed unmixing matrix from ICA. Substituting for X_{red} from eqn. (4.2) gives:

$$C_{red} = WV^T X \tag{4.4}$$

To estimate the timecourses of the maps C_{red} calculated from the dimension-reduced data, we wish to find a matrix T, such that

$$X \approx TC_{red} \tag{4.5}$$

An estimate for T can easily be obtained by scaling the rows of C_{red} so that $|C^i_{red}| = 1$. Thus, $C_{red}C^T_{red} \approx I$, because the rows of C_{red} are maximally independent. Finding the p time courses (of length n) associated with each of the p maps can now be determined by examining the columns of the matrix,

$$T \approx XC^T_{red} \tag{4.6}$$

In this paper, we use PCA & ICA (McKeown et al., 1998; McKeown and Sejnowski, 1998) to separate fMRI data from an event-related study with a human subject into spatially independent components and their associated timecourses to indirectly infer the amount of randomness in the data, as assumed by typical analysis of event-related data.

4.2 Methods

We performed a simple event-related fMRI visual experiment with a single normal subject. Ninety-five visual stimuli, consisting of black and white radial checkerboards that subtended about 20^o by 15^o of visual angle and were presented singly for 500ms. The interval between successive stimuli varied randomly between 14 to 18s. Ten contiguous 5mm thick slices were acquired parallel to the line connecting the anterior and posterior commissures (axial imaging plane) on a 1.5T machine. Functional gradient echo echo-planar images were acquired at a TR of 1s (TE: 40ms, Flip Angle: 81^o, FOV: $24cm$, matrix: 64^2, in-plane resolution: $3.75mm^2$).

The fMRI data were temporally aligned to compensate for interleaved slice acquisition. Because we were investigating responses of the sluggish BOLD signals to individual stimuli, the data were low-pass filtered with a Hanning window of length 4 to increase the signal-to-noise ratio (Press et al., 1992). The raw data were not spatially smoothed. A minimum-intensity threshold was used to identify voxels in the head.

The data set was reduced in dimension using Principal Component Analysis (PCA – (Jackson, 1991)) to one hundred components. Independent Component Analysis (ICA) was applied to the dimension-reduced data set, separating the data into 100 spatially-independent components (McKeown et al., 1998; McKeown and Sejnowski, 1998).

The Bell-Sejnowski algorithm for ICA is an iterative technique that starts with an initial weight matrix that is iteratively updated until convergence is reached. Typically the initial weight matrix is set to the identity matrix. To determine which components were robust to weight matrix initialization, the data were re-separated, except this time the initial weight matrix was the inverse of the calculated weight matrix from the first separation.

As the ICA algorithm does not assign any direct importance to the order of the components, the different sets of spatially-independent components from the two data separations were compared for similarity. Component pairs were first created by selecting a component from one group of separated components and selecting another component from the other group of separated components. When two component sets were compared, the mutual information between all possible pair combinations was computed to determine the components that matched the most closely. This was accomplished by separating the first component values into 20 histogram bins. The voxels in the second component that corresponded to those voxels from the first component contained in the first histogram bin were

also separated into 20 histogram bins. This process was repeated until a 20-by-20 contingency table of the data was created. The computations were normalized so that a component compared to itself would result in a value of 1, and two completely random components would result in a value of about 0, see eqn. 14.4.17 in (Press et al., 1992). The same component was not allowed to contribute to more than one component pair.

The mutual information between best-matching component pairs was plotted to determine which components were robust to weight matrix initialization. A cutoff for the mutual information, based on the inflection point of the mutual information vs. component number curve was estimated by visual inspection. Component pairs that matched above the cutoff were deemed reproducible and deterministic, and component pairs whose mutual information was below the cutoff were deemed unreproducible and assumed to refer to random noise. Specifically, the data were modeled as,

$$X = T \cdot C = [T_{rep}|T_{noise}] \left[\frac{C_{rep}}{C_{noise}} \right] \tag{4.7}$$

$$T = [T_{rep}|T_{noise}] = C^T X \tag{4.8}$$

where X is the all time-point data set, of dimension n by v, where n is the number of time points and v is the number of brain voxels in all slices. The T and C matrices were partitioned to correspond to the reproducible and unreproducible components defined by the cutoff described above, of dimensions n by 100 and 100 by v respectively.

In order to assess the relative contribution to the entire data of the reproducible components,

$$X_{rep} = [\ T_{rep}|0] \ \left[\frac{C_{rep}}{0} \right] \tag{4.9}$$

was calculated. The relative variance in the data that was deterministic was estimated as:

$$Ratio_{Det} = \frac{tr(cov(X_{rep}^T))}{tr(cov(X^T))} \tag{4.10}$$

where $tr()$ refers to the trace of a matrix, and cov() refers to the covariance matrix.

The reproducible components were then examined to see if they had a relationship to the stimulus presentation. The time course of each reproducible component was divided into epochs by taking the 5 time points preceding to 13 time points following checkerboard onset. Each 19 time point epoch was linearly detrended. The mean of each component across epochs was calculated to create a mean vector, of 19 time points, M. The entire time course of the component was then correlated with the convolution of M and a vector containing ones at the times of stimulus presentation. The degree of correlation was assumed to represent the amount that component was time-locked to stimulus presentation. To estimate the significance of the correlation,

under the null hypothesis that the two waveforms were not correlated, we calculated

$$p = erfc\frac{|r|\sqrt{n}}{\sqrt{2}} \tag{4.11}$$

where n is the number of time points, and $|r|$ is the magnitude of the correlation coefficient (see also Press et al., 1992). Components that were significant to $p < 0.001$ were selected as being affected by stimulus presentation.

The five most significantly stimulus correlated components were overlaid on an anatomical image obtained at the same sitting. The spatial maps of the ICA components were first spatially smoothed with a $6mm$ fwhm (full-width-half-maximum) Gaussian filter. The smoothed maps were thresholded at $z > 1.25$.

4.3 Results

The fMRI series created a data set consisting of $1,464$ time points by $10,192$ brain voxels. The plots of mutual information between component pairs are shown in figure 4.1.

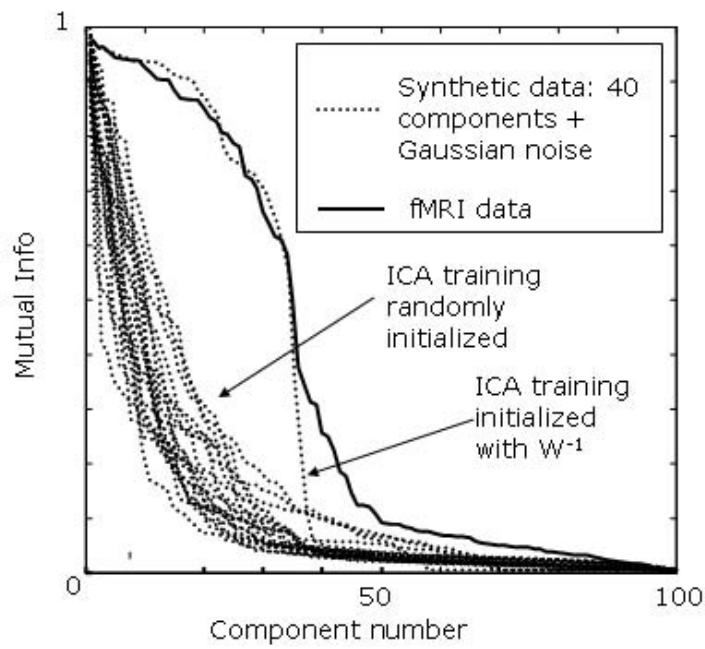


Figure 4.1 The plots of mutual information between component pairs. The mutual information between best matching component pairs (without duplication of any single component) is shown after two ICA separations with different initial weight matrices. Note the abrupt inflexion point at MI values of ~ 0.1.

Best matching pairs had a mutual information of about 0.9 with least matching pairs having a mutual information of < 0.05, suggesting almost complete statistical independence. To estimate the mutual information expected by chance, we plotted a histogram of all elements of the 100 by 100 matrix whose elements, M_{ij}, referred to the mutual information between the i^{th} component from the first separation and the j^{th} component of the second separation. This suggested that almost all component pairs had mutual information measures of ≤ 0.1 (figure 4.2).

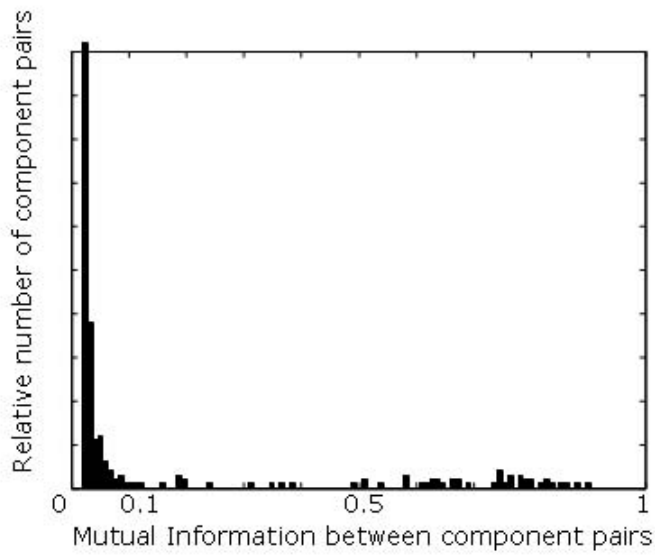


Figure 4.2 To estimate the mutual information expected by chance, we plotted a histogram of all elements of the 100 by 100 matrix whose elements, M_{ij}, referred to the mutual information between the i^{th} component from the first separation and the j^{th} component of the second separation. This suggested that almost all component pairs had mutual information measures of ≤ 0.1.

A relatively abrupt change in the slope of the Mutual Information curve was evident on visual inspection at $n = 55$. This value was used as the arbitrary cutoff to differentiate between reproducible or deterministic features of the data, and non-reproducible or random features (cf. equations 9 and 10). The ratio of deterministic variance to total variance (equation 10) was 0.84. Of the 55 reproducible components, 28 independent components were considered correlated with the stimulus presentation (equation 11) to a significance of $p < 0.001$.

To test whether the weighting the second ICA separation with the inverse of the weight matrix created an uncontrollable dependency between the repetitions, we performed simulations on synthetic data. An artificial data set consisting of 40 ICs corrupted with pure Gaussian noise was constructed. The first 40 spatial ICs from one ICA separation were pre-multiplied by a random 100×40 matrix (each element drawn from uniform distribution [0 1]).

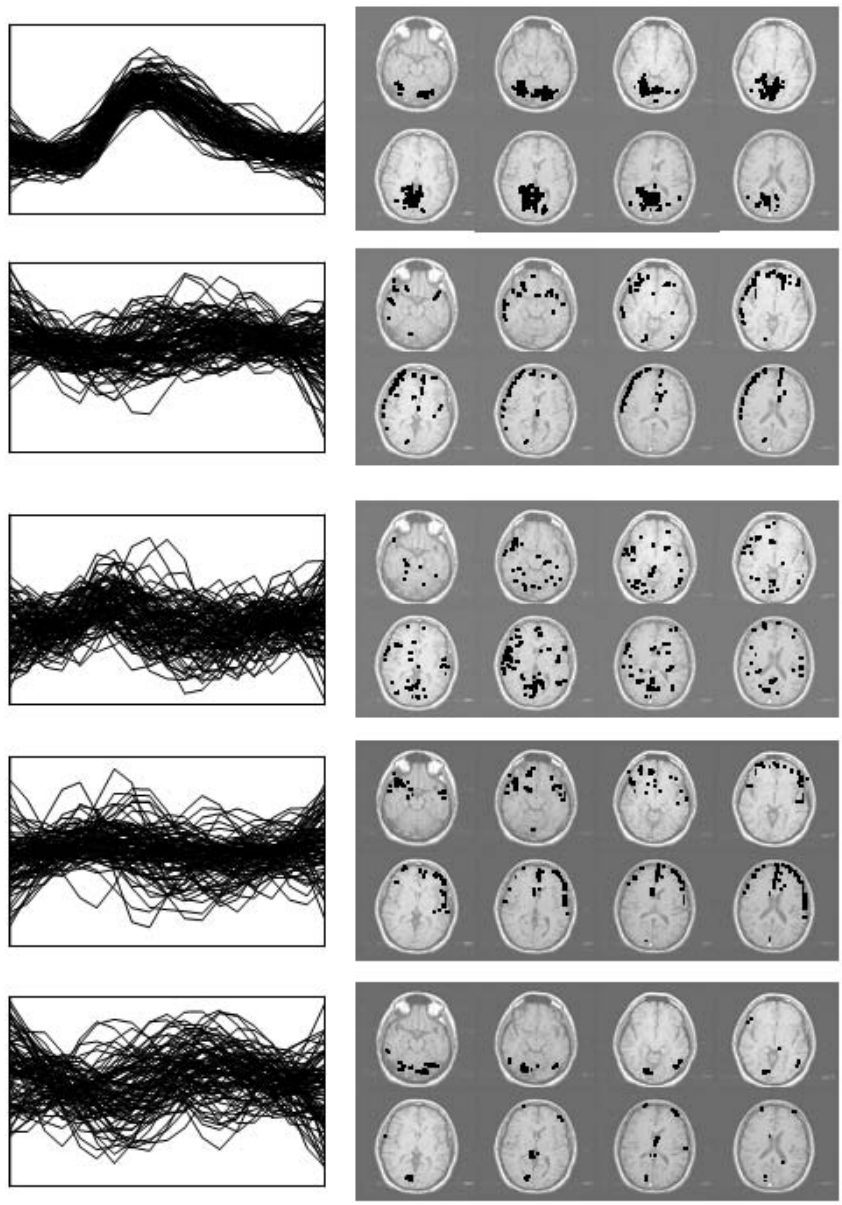


Figure 4.3 The five most significantly stimulus-correlated (all $p < 10^{-19}$) components. (*column* 1) The time courses of these components were divided into 19 timepoint epochs time-locked to stimulus presentation, linearly detrended and overlaid. (*a*) The Most Significantly Correlated (MSC) component. The time course of this component most closely followed that expected from a single visual flash. (*b*) & (*d*) The ring-like spatial structures of these components suggest movement, but their time courses seem reliably reproduced by the task. (*c*) Task-related brain activation loading heavily in temporal regions. Contrast the relative onset of this component to (*a*). (*e*) Occipital task-related component with a different temporal profile. Note the relative "jitter" between events for this component compared to MSC component in (*a*).

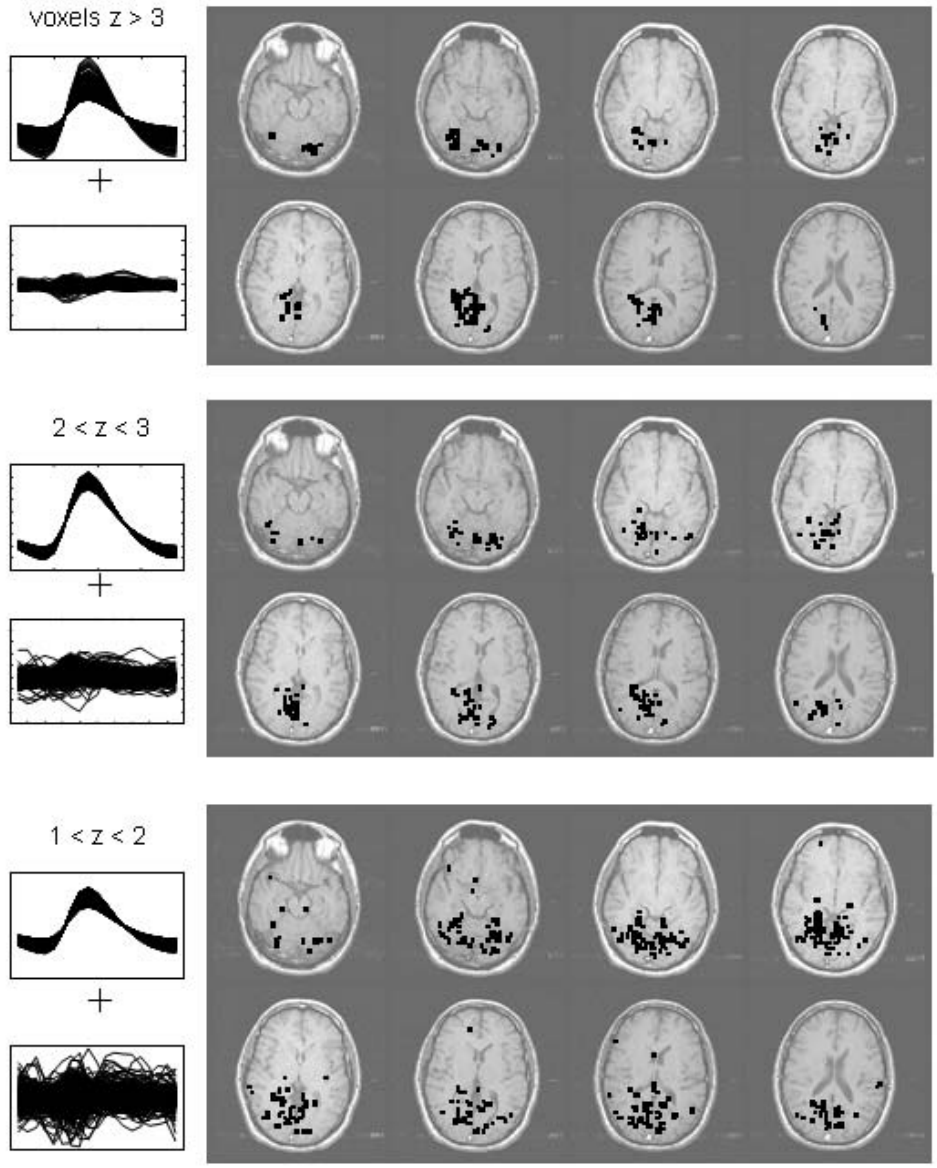


Figure 4.4 Relative contribution of different components in selected voxels. Voxels loading heavily in the MSC component (4.3a) were examined to determine the relative contribution of different components to these voxels. *(top)* Voxels loading most heavily in the MSC component *(top, right)* received contributions to their raw time courses from the MSC component and other components *(top left, same vertical scale shown for both figures)*. Each trace represents a different voxel. *(middle, bottom)* Voxels loading less heavily on the MSC component received greater and greater relative contribution from other components *(middle and bottom left, same vertical scale shown for both figures)*.

The resultant matrix was made full rank by adding Gaussian noise, with zero mean and unit standard deviation. Two initial ICA separations were performed as before, with the second initialized with the inverse of the first weight matrix. Also, a comparison between ICs initialized with the identity matrix was compared with many repetitions (30) of ICs initialized with a random matrix. The seeding with the inverse weight matrix appeared to more accurately reflect the true number of components in the synthetic data (i.e., 40) (figure 4.1).

The five most significantly stimulus-correlated (all $p < 10^{-19}$) components are shown in figure 4.3. The most significantly stimulus-correlated ICA component demonstrated very little trial-to-trial variability and appeared directly task-related (figure 4.3a). The spatial distribution of this component (thresholded at $z \geq 1.25$) was heavily weighted in occipital regions. At least two significant components were presumably the result of task-dependent head motion (figure 4.3b,d), a phenomenon previously described (Bullmore et al., 1999; Rombouts et al., 1998; Thacker et al., 1999). Other significant components were weighted heavily in occipital regions (overlapping with the component in 4.3a), with distinct, but still stimulus-dependent temporal profiles (figure 4.3c,e). One of these components appeared to involve temporal and frontal regions of brain activation (figure 4.3c).

To see the relative effects that these stimulus-correlated components might have, we examined voxels that weighted heavily on the most stimulus-correlated (MSC) component (figure 4.3a). We then examined the relative contributions of this component (by taking the outer product of its time course and spatial map) and all other components at these voxels (figure 4.4). The spatial map associated with the MSC component was converted to z-scores. As expected, in voxels loading most strongly on this component ($z \geq 3$), the MSC component was the main contributor to the raw time course of these voxels (figure 4.4, *same scale used for MSC component contribution and contribution from all other components*). However, on voxels loading less well, for example, $1 < z < 2$, contribution from other components were of the same order of amplitude as the contribution from the MSC component (figure 4.4, bottom).

4.4 Discussion

Our results suggest that the majority of the variance of fMRI data can be effectively modeled as robust spatially independent components. We found a number spatially independent components whose mean time course remained significantly correlated with the empirical waveform when averaged over a reasonable number ($n = 95$) of trials (figure 4.3). We refer to these components as "stimulus correlated" because their mean, when averaged time-locked to stimulus presentation appeared significantly correlated with the presentation of stimuli. The most significantly stimulus-correlated component appeared directly task related and was temporally and spatially similar to that expected from prior studies of averaging the raw data (figure 4.3a). However other stimulus correlated components appeared related to motion

induced artifact (figure 4.3b,d), overlapping occipital regions with differing temporal profiles (figure 4.3e), and frontal and temporal regions of unknown significance (figure 4.3c).

The presence of different stimulus-correlated components that were suggestive of true brain activation and had temporal profiles distinct from the MSC component (figure 4.3c,e) is consistent with the notion that fMRI data consists of temporally and spatially overlapping components. Moreover, these different stimulus-correlated components appear to have slightly different latencies from event-to-event, indicating the need to align each component separately as has recently been suggested for ERP literature (Jung et al., 2000). Of the stimulus-correlated components, clearly not all were related to cortical processing, as at least two were related to stimulus correlated movement (figure 4.3b,d). However, isolation of stimulus correlated components is still important, as their effects will tend to become more pronounced when data are averaged time locked to stimulus presentation, as is typically done with event-related fMRI studies.

An important result from any fMRI experiment is the extent of activation. In voxels that loaded most heavily in the MSC component (figure 4.4 top), the task-related activity appeared to dominate the raw signal, so averaging the raw data would not affect the interpretation that these voxels are indeed task-related. However, in voxels that loaded less heavily, other task-related components and non-task related components that were averaged time-locked to stimulus presentation were of the same order of magnitude as the task-related signal (figure 4.4 bottom). These voxels were still contiguous and loaded heavily in occipital regions suggesting that they represent true activation in response to visual stimuli. Averaging the raw data in this case may result in erroneous interpretations as to the true extent of activation.

We note that the components themselves exhibited much less trial-to-trial variability than a typical stimulus correlated voxel (not shown). This suggests that the variability seen in an individual voxel is more the effect of spatially-overlapping deterministic processes than of truly random noise, and suggests a role for deterministic models (such as ICA) to isolate the stimulus correlated changes at a voxel. This may ultimately lead to a reduced number of total trials that may need to be averaged before a stable value is reached. Further work needs to be done to directly compare the number of trials required to stabilize the spatial and temporal estimates of the individual stimulus correlated components as opposed to averaging the raw data.

Our results suggest that averaging raw fMRI data to infer the temporal profile and spatial extent of activation must be done with caution. Either a restricted region of interest (ROI) could be used, or first techniques like ICA can be used to segregate voxels with similar properties before averaging.

References

P. A. Bandettini and R. W. Cox. Event-related fmri contrast when using constant interstimulus interval: theory and experiment. *Magnetic Resonance in Medicine*, 43(4):540–8, 2000.

G. M. Boynton, S. A. Engel, G. H. Glover, and D. J. Heeger. Linear systems analysis of functional magnetic resonance imaging in human v1. *Journal of Neuroscience*, 16(13):4207–21, 1996.

R. L. Buckner, P. A. Bandettini, K. M. O'Craven, R. L. Savoy, S. E. Petersen, M. E. Raichle, and B. R. Rosen. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging [see comments]. *Proceedings of the National Academy of Sciences of the United States of America*, 93(25):14878–83, 1996.

R. L. Buckner, W. Koutstaal, D. L. Schacter, A. M. Dale, M. Rotte, and B. R. Rosen. Functional-anatomic study of episodic retrieval. ii. selective averaging of event-related fmri trials to test the retrieval success hypothesis. *Neuroimage*, 7 (3):163–75, 1998.

E. T. Bullmore, M. J. Brammer, S. Rabe-Hesketh, V. A. Curtis, R. G. Morris, S. C. Williams, T. Sharma, and P. K. McGuire. Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fmri. *Human Brain Mapping*, 7(1):38–48, 1999.

M. A. Burock, R. L. Buckner, M. G. Woldorff, B. R. Rosen, and A. M. Dale. Randomized event-related experimental designs allow for extremely rapid presentation rates using functional mri. *Neuroreport*, 9(16):3735–9, 1998.

K.J. Friston. Statistical parametric mapping and other analyses of functional imaging data. In A.W. Toga and J.C. Mazziotta, editors, *Brain Mapping, The Methods*, pages 363–396. Academic Press, San Diego, 1st edition, 1996.

A. M. Howseman, D. A. Porter, C. Hutton, O. Josephs, and R. Turner. Blood oxygenation level dependent signal time courses during prolonged visual stimulation. *Magnetic Resonance Imaging*, 16(1):1–11, 1998.

S. A. Huettel, J. D. Singerman, and G. McCarthy. The effects of aging upon the hemodynamic response measured by functional mri. *Neuroimage*, 13(1):161–75, 2001.

J. E. Jackson. *A Users Guide to Principal Components*. John Wiley & Sons, Inc., New York, 1991.

T-P Jung, S. Makeig, T-W Lee, M.J. McKeown, G. Brown, A.J. Bell, and T.J. Sejnowski. Independent component analysis of biomedical signals. In *The 2nd Int'l Workshop on Independent Component Analysis and Signal Separation*, pages 633–44, 2000.

G. McCarthy. Event-related potentials and functional mri: a comparison of localization in sensory, perceptual and cognitive tasks. *Electroencephalography & Clinical Neurophysiology - Supplement*, 49:3–12, 1999. Review.

G. McCarthy, M. Luby, J.C. Gore, and P. Goldman-Rakic. Functional magnetic resonance imaging in a visual oddball task. *NeuroImage*, 3:S548, 1996.

M.J. McKeown, S. Makeig, G. Brown, S. Kindermann, and T.J. Sejnowski. Modes or models: a critique on independent component analysis for fmri - reply to commentary by prof friston. *Trends in Cognitive Sciences*, 2:375, 1998.

M.J. McKeown and T.J. Sejnowski. Independent component analysis of fmri data: Examining the assumptions. *Human Brain Mapping*, 6:368–372, 1998.

W. H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 2nd edition, 1992.

S. A. Rombouts, F. Barkhof, F. G. Hoogenraad, M. Sprenger, and P. Scheltens. Within-subject reproducibility of visual activation patterns with functional magnetic resonance imaging using multislice echo planar imaging. *Magnetic Resonance Imaging*, 16(2):105–13, 1998.

B. R. Rosen, R. L. Buckner, and A. M. Dale. Event-related functional mri: past, present, and future. *Proceedings of the National Academy of Sciences of the United States of America*, 95(3):773–80, 1998. Review.

N. A. Thacker, E. Burton, A. J. Lacey, and A. Jackson. The effects of motion on parametric fmri analysis techniques. *Physiological Measurement*, 20(3):251–63, 1999.

A. L. Vazquez and D. C. Noll. Nonlinear aspects of the bold response in functional mri. *Neuroimage*, 7(2):108–18, 1998.

5 Exploratory Analysis of Event-Related fMRI Demonstrated in a Working Memory Study

Andrzej Wichert, Henrik Walter, Jo Grothe, and Birgit Abler
Department of Psychiatry
University of Ulm, D-89075 Ulm, Germany
wichert@neuro.informatik.uni-ulm.de
{henrik.walter, jo.grothe, birgit.abler}@medizin.uni-ulm.de

Friedrich T. Sommer
Department of Neural Information Processing
University of Ulm, D-89069 Ulm, Germany
fritz@neuro.informatik.uni-ulm.de

Here we describe a novel technique for exploratory analysis of event-related fMRI. The technique comprises two parts. The first component is dense latency sampling (DLS), an oversampling scheme for event-related fMRI that has the advantage of providing volume slice timing without the need for signal interpolation. The second component is dynamical cluster analysis (DCA) of signal time-courses; this analysis is done with temporal constraints taken from the event-related design: Signal segments that correspond to different types of events are analyzed separately to reveal specific event-related activation. The technique does not rely on preassumptions about the temporal shape of functional activity like common inerential methods.

We demonstrate the utility of the technique and also compare its performance to standard techniques in a study of working memory. The technique reveals spatio-temporal patterns of activity associated with different memory load conditions. Most delay-related activity appeared in parietal and prefrontal regions peaked in the second half of the delay period; this suggested involvement of these regions in processes of memory rehearsal and decision making. The superior parietal and precentral cortices also participated in delay-related activity. But for these regions the temporal shapes of functional activation suggested additional roles in memory encoding.

5.1 Introduction

Almost all methods of data analysis in fMRI make assumptions about the nature of the underlying neuronal processes. The methods can be divided into two classes, based on the type of assumptions made. One class of methods uses univariate analysis and relies on the assumption of functional specialization of cortical regions. The other class employs multivariate analysis and relies on the assumption of functional integration, that is, that brain function results from cooperative interactions among cortical regions. Here, we will describe a technique, belonging to this second class, that applies multivariate analysis to data from event-related fMRI (ER fMRI). Our approach of exploratory data analysis is designed to detect weak, task-induced signal changes whose shapes are not known beforehand (Wichert et al., 2001a,b). The technique combines a new scheduling scheme of multi-slice data acquisition with a variant of temporal cluster analysis. The approach is designed for complex experimental paradigms with short events where the event-related signal is weak and has a time course that is difficult to predict. Studies that explore cognitive processes like memory typically involve such complex paradigms.

In order to evaluate the strength of our technique we chose the study of working memory. The choice of this cognitive task made sense for two different reasons. First, working memory is amenable to study with our technique because the timing of the neural processes it involves can be directly manipulated by experimental design. Second, earlier studies had indicated that the formation of working memory is distributed across disparate cortical areas, thus, suggesting that functional integration might be important.

The concept of working (or short term) memory refers to a type of memory that has limited capacity for storing and manipulating information necessary for performing a specified task. It was originally defined in studies in which subjects were presented with a list of items and then asked to recall individual entries on the list. Results from these studies defined the upper limits of memory load, i.e., the maximum number of items (such as words) that could be memorized with reasonable accuracy. Typically, experiments test working memorizes whose durations range from a few seconds to a few minutes. The mechanisms of working memory vary as a function of the length of time the memory is required to last. If the duration is short, 10s or less, subjects recall items nearer the end of the list more accurately that those at the beginning of the list. This recency effect disappears if the duration of the memory exceeds 10s. Thus, there seems to be a qualitative difference between memories that persist for more 10s and those that are briefer — the former are more durable than the latter.

A common type of task used in studies of working memory is called delayed match-to-sample task or simply delayed response task. The task consists of three discrete phases. The first is a presentation phase during which the subject is presented with a set of items to memorize (the memory set). The presentation phase is followed by the delay period, an interval during which no tasks are required. The

last phase is the probe phase during which the subject is presented with an item and must decide whether or not it belongs to the memory set.

The cellular neurophysiology of delayed response tasks has been studied in experiments with animal models. These studies showed that prefrontal and parietal cortical region are involved in working memory. Cells in these regions respond selectively during different phases of the task, indicating "process specificity" (Baddeley, 1986, 1996). For instance, a population of cells in prefrontal cortex fire persistently during the delay phase. Cells in prefrontal cortex are also able to convey information about both the identity and location of a given item in a memory set (Rainer et al., 1998).

Whole-brain neuroimaging techniques such as PET and fMRI promise to reveal the global functional architecture of working memory (Jonides et al., 1993; D'Esposito et al., 1995; Goldman-Rakic, 1996; Owen et al., 1996; Postle et al., 2000b; Goldman-Rakic, 2000). The first neuroimaging studies of working memory made only indirect assessments of functional activity (Cabeza and Nyberg, 2000). The advent of the event-related fMRI technique (Josephs et al., 1997; Dale and Buckner, 1997) gave direct access to the functional activity caused by short events such as a single delayed response task. These later studies led to revisions of theories of working memory that had been based on results from the indirect assessments (Postle et al., 2000a). Thus, the evolution of methods in whole-brain imaging allows the refinement and revisions of theories of brain function.

In the study of working memory that we will describe we will focus on the following three questions. i) How do results depend on the paradigm used in data analysis, specifically, do results change if one switches from the assumption of functional specialization (underlying common univariate inferential fMRI data analysis) to the assumption of functional integration (the incentive of multivariate exploratory fMRI data analysis)? ii) How do the time courses of functional activity relate to results from single-cell recordings? iii) Are sensory areas involved in the delay phase?

All told, the overall aims of the chapter are to explain our data acquisition and analysis technique, to discuss its relations to other approaches in ER fMRI, and to evaluate its ability to explore processes of working memory. The results of the new technique will be compared with those of state-of-the-art approaches, i.e., the slice timing technique usually applied in multi-slice fMRI, and conventional inferential data analysis by the general linear model (GLM). The discussion of the results of our study in the context of earlier working memory literature will be brief and certainly not cover all aspects. A more exhaustive description of the results will appear in a forthcoming paper.

5.2 Current Methods for Event-Related fMRI

5.2.1 New Chances and Challenges

The classical block design common in PET and early fMRI introduced stationary phases of functional activation (blocks). The stationarity requirement posed strong limits on the investigation of behavioral paradigms. A considerable widening of the scope of neuroimaging was provided by the introduction of ER fMRI registration technique. In ER fMRI the data aquisition is exactly scheduled relative to events in the experimental design. Thus, it allows to register signal changes evoked by short events, quite similar to evoked event-related potentials measured with EEG/MEG. Of course, with regional blood flow imaging techniques the temporal resolution is generically limited by the delayed and low-pass filtered hemodynamic response (HR) (with time-to-peak interval of about 5s). However, up to the generical resolution limit, ER fMRI provides more freedom for implementing experimental paradigms. For instance, short events can be repeated in random order, or categorized post-hoc. Early studies using the event-related technique were on odd-ball paradigms (MCarthy et al., 1997) and on various cognitive paradigms (for short comprehensive reviews, see (Buckner, 1998; Rosen et al., 1998)). In the performance of cognitive tasks, ER fMRI allows to discern and characterize different phases that were only indirectly assessible in block designs.

The extended scope of ER fMRI implicates as well new difficulties in data analysis. For inferential analysis the availability of an adaequate regressor or model function for functional activation, determined external to the data, is an indispensable prerequisite (Lange, 1996, 1997; Lange et al., 1999; Petersson et al., 1999a,b). The regressors used for block designs are box-car functions reflecting the blocks, convolved by a canonical HR (Bandettini et al., 1993). If the durations of blocks are long compared to the HR characteristic, exact modeling of the HR is uncritical for the inferential analysis. For event-related designs the situation is entirely different. The time course of functional activity is not as completely prescribed by the experimental paradigm as in block design. Thus adaequate regressor functions necessary to detect the weak component of functional activity[1] are hard to predict independently of the fMRI data. A method to estimate adaequate regressor functions from the fMRI data set is revisited in the following paragraph.

1. The typical S/N of functional imaging in an 1.5T scanner is 1-5%. Regression analysis with the box-car shaped regressor function reflecting the block design reduces the noise by temporal averaging within blocks.

5.2.2 Signal Averaging and Data Analysis

To achieve a noise reduction in ER fMRI *selective averaging* has been proposed by Dale and Buckner (1997):

$$h = X^T y \tag{5.1}$$

where h is the sampled HR, y stands for the signal measured at a voxel, and X is the design matrix reflecting the event timing during the experiment.

Current inferential data analyses for ER fMRI uses families of regressor functions generated by a canonical HR impulse response function systematically shifted in time. Since for short events the sensitivity of the analysis depends critically on a good model of the HR, Dale (1999) had proposed to estimate the HR response from the data set to be analyzed. He used univariate linear signal estimation based on ordinary least square (OLS) fits

$$h_{OLS} = (X^T X)^{-1} X^T y, \tag{5.2}$$

To take into account influences between events, Burock and Dale (2000) employed univariate linear signal estimation based on maximum likelihood (ML) estimation:

$$h_{ML} = (X^T C_n^{-1} X)^{-1} C_n^{-1} X^T y. \tag{5.3}$$

where C_n denotes the covariance of the noise. Burock and Dale (2000) used equation 5.3 as regressor in a modified approach of inferential data analysis.

5.2.3 Increased Resolution by Oversampling

In state-of-the-art fMRI scanning the lower limit on the repetition time (TR) is methodologically prescribed. Since different slices cannot be acquired simultaneously, TR grows proportionally with the number of slices in the measurement volume. Because the time constant of spin relaxation is fixed, the time required for a slice measurement cannot be arbitrarily reduced. The current limit is about $TR \simeq k \times 80$ ms, with k the number of slices. As first suggested for cardiac fMRI, oversampling can virtually increase the intrinsically low temporal resolution of multi-slice ER fMRI[2]. Oversampling means that each type of event is recorded repeatedly, say r times, each event repetition sampled with a different latency. The sampling time points can either be randomly jittered, or equidistantly distributed in the interval TR. The latter we refer to as *equidistant oversampling*. In such sampling schemes the latency is varied between 0 and $TR(1 - 1/r)$ in steps of TR/r which increases the effective sampling rate from $1/TR$ to r/TR. Josephs et al. (1997) first proposed equidistant oversampling for neuro fMRI (with $r = 2$). The method was used for estimating HR functions (Miezin et al., 2000).

2. Oversampling critically relies upon the condition that repeated trials produce similar functional activation, an assumption also made in conventional fMRI.

5.2.4 Slice-Timing

The technical limitation in multi-slice fMRI that different slices cannot be recorded simultaneously causes the *slice timing problem*. This describes the fact that a recorded volume is not an instantaneous picture in time. The different slices are recorded one after another in intervals of TR/k. While negligible with traditional block designs, the slice timing differences matter for the investigation of event-related designs.

The current method to solve this problem is a procedure called volume slice timing that involves phase-shift manipulations in the data: A reference slice is selected in the volume and all signals measured in other slices are phase-shifted[3] to the sampling points of the reference slice, see figure 5.1 a). Because the required shifts of the phases are smaller than the sampling intervals of the signals, this manipulation involves signal interpolation and the typical errors associated with it, such as ringing and wrap-around effects, see (Schanze, 1995). The described procedure is applied as a standard preprocessing step, often even before inferential data analysis, where in principle the slice timing could be reflected more properly by shifting the regressors for each slice individually (Josephs et al., 1997).

5.3 Exploratory Analysis in Functional Imaging

For multivariate exploratory analysis of fMRI/PET data various methods have been proposed, such as principal component analysis (Lai and Fang, 1999; Hansen et al., 1999; Baumgartner et al., 2000a), indepedent component analysis (McKeown et al., 1998; McKeown and Sejnowski, 1998; McKeown et al., 1999), and diverse temporal clustering methods (Scarth et al., 1995; Baumgartner et al., 1997; Golay et al., 1998; Baune et al., 1999; Goutte et al., 1999; Filzmoser et al., 1999; Fadili et al., 2000), and see chapter Somorjai and Jamusz. The goal of these approaches is to detect characteristic spatio-temporal properties in the data as much as possible uninformed of a priori assumptions about the results.

Exploratory data can yield a reduced data set that still reflects the important properties in the data. For instance, cluster analysis approaches in functional imaging usually apply *temporal cluster analysis* (TCA) i.e., they cluster the data with respect to the shapes of the signal time courses. TCA partitions the data into sets of voxels with similar time courses—clusters[4]. A cluster can be characterized by its *spatial pattern* and by the *cluster center*, that is, the averaged time course.

3. Applying the Fourier-shift theorem, a phase shifting can be achieved by multiplying with a complex exponential in Fourier space (Aguirre et al., 1998).
4. It has to be emphasized that a cluster resulting from temporal TCA is different from a spatially contiguent set of voxels (for instance, in functional maps), often referred to as cluster. The first collates voxels of similar signal shape, but completely independent of spatial positions. To avoid confusion we will refer to the latter as a spatial cluster.

Exploratory data analysis was successfully applied for block design experiments, see the comparison between different exploratory and inferential data analysis approaches in (Lange et al., 1999). However, few attempts have been made to apply exploratory data analysis to ER fMRI, but see (Richter et al., 2000). For ER fMRI, where inferential data anlysis is hampered by the lack of adaequate regressors, exploratory data is particularly interesting.

5.4 New Technique for Exploratory Analysis of Event-Related fMRI

In this section we describe a new approach to characterize functional activity in event-related fMRI. It relies on selective signal averaging as well, but applies exploratory analysis techniques (section 5.3) rather than univariate signal estimation and inferential data analysis described in section 5.2.2. The first part of this section, section 5.4.1, explains the data acquisition method, the second part, section 5.4.2, the exploratory analysis algorithm and its application to ER fMRI data.

5.4.1 Data Acquisition and Slice Timing

5.4.1.1 Dense Latency Sampling

For increased temporal resolution we use equidistant oversampling of events as described in section 5.2.3. Applying selective averaging (equation 5.1) after equidistant oversampling yields signal time courses virtually sampled with a rate of r/TR.

Equidistant oversampling can have another interesting consequence for multi-slice data acquisition that, as far as the authors are aware, has not been exploited before: With the sampling rate chosen appropriately it can also resolve the slice timing problem described in section 5.2.4; for $r = k$, that is the number of repeated events is equal to the number of acquired slices, a dense sampling can be guaranteed. Dense sampling means that for any slice measurement all the $k-1$ other slices completing a volume have been recorded with the same latency during other repetitions of the event. Thus, slice timing can be achieved just by data re-sorting, i.e., by rearranging slices of same acquisition latency to new volumes. Re-sorting is done with respect to a labeling of the measured data based on the event-related design matrix. Each slice is labeled by the latency between the exact acquisition time and the event with the largest influence on the signal. The labeling takes into account an assumed delay until the maximum effect of an event is expressed in the HR. We used a 5s time-to-peak interval of the canonical HR. The described methods are explained by schematic pictures in figure 5.1.

The combination of equidistant dense sampling scheme and data re-sorting we call the *dense latency sampling (DLS)* technique (Wichert et al., 2001b). Like in conventional ER fMRI the DLS technique leaves the freedom to schedule order and onsets of events in the experiment in a pseudorandom manner. The advantage of the DLS fMRI technique is that it provides volume slice timing without introducing

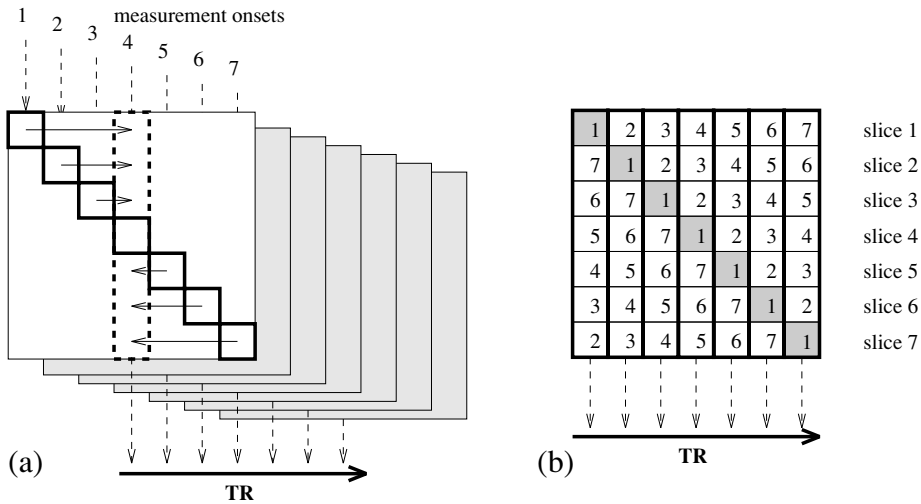


Figure 5.1 Schematic views of the DLS method and phase-shift slice timing (example with $k = 7$ slices per volume and $r = 7$ event repetitions). Picture a) shows the DLS data acquisition scheme and the effect of phase-shift slice timing: Big squares symbolize the acquisition of volumes. The horizontal extension corresponds to the acquisition time TR, the vertical extension to the spatial axis perpendicular to the measured slices (as labeled on the right margin of the figure). The shaded squares in the background sketch volume measurements taken at event repetitions. Latencies between events and measurements were varied such that the interval of TR (marked by the bold horizontal arrow) is equidistantly divided by sampling points (depicted by the downward arrows). Thus, the oversampling rate is TR/r. Within the acquisition volume in the foreground the small squares symbolize how different slices are measured one after the other. The thin horizontal arrows depict the effect of phase-shift slice timing (with slice 4 chosen as the reference slice), the dashed rectangle symbolizes the resulting slice-timed volume. Picture b) shows the result of the DLS-re-sorting. Bold rectangles denote the volumes assembled by re-sorting with respect to the latencies between event and the individual measurements. The numbers indicate the event repetition from picture a) indicating the origin of each slice. In addition, the slices acquired during the first event are marked by shaded squares—they are distributed over different DLS volumes. Note, that the choice $r = k$, provides complete volumes at each sampling point after re-sorting.

artifacts of the conventional slice-timing procedure (section 5.2.4). Particularly for designs with long repetition times TR, when phase-shift artifacts become more and more serious, the DLS technique offers an interesting alternative for slice timing. A reliable method for volume slice timing is crucial for all sorts of exploratory analysis methods such as principal/independent component analysis and cluster analysis.

5.4.1.2 Dense Latency Sampling at Lower Rates

As explained before, the full DLS technique requires an oversampling rate TR/r equal to the single-slice sampling rate of the used measurement sequence. For complex experimental paradigms, however, this requirement results in long durations

of experimental sessions. In such cases one might prefer a lower effective temporal resolution, if achieved with fewer event repetitions.

A reduction of the sampling rate required for data re-sorting is possible by combining DLS-fMRI with modest phase-shifting. The idea is to apply phase-shift time slicing not only with respect to a single reference slice per volume. Instead one can choose $s > 1$ reference slices spaced equidistantly over the volume. For each slice only the data corresponding to the closest reference slice are used. These can be arranged to complete volumes by DLS re-sorting as explained in section 5.4.1.1. Compared with the traditional phase-shift slice timing the resulting interpolation errors are smaller because the maximum phase-shifts involved are reduced from $TR/2$ to $TR/(2s)$[5]. The combination of DLS-technique and phase-shifting can be best explained with an example: For a measurement volume consisting of $k = 21$ slices, phase shifting is done for the reference slices $2, 5, 8, 11, 14, 17$, and 20. The result is $s = 7$ slice-timed zones in the volume, each zone comprising three adjacent slices. The situation is again reflected by figure 5.1 where now each slice corresponds to a zone of 3 slices and the measurement latencies stand for the latencies of the reference slices. By DLS-re-sorting one can rearrange the different zones to complete slice-timed volumes. In this example the maximum phase-shift to be applied to the data is reduced from $10 \times TR/21$ in the standard slice timing technique (using slice 10 as reference slice) to $1 \times TR/21$ with the DLS technique. Thus, the required event repetitions, as well as the effective sampling rate were reduced by a factor of 3 compared with the full DLS-scheme.

5.4.2 Exploratory Analysis for Event-Related fMRI

5.4.2.1 Restricted Cluster Analysis

The results of multivariate exploratory analyses like temporal cluster analysis (TCA) depend on the selection of voxels included in the analysis. Therefore, removal of voxels outside the brain using a simple threshold criterion for the mean signal amplitude is a usual preprocessing step. The most assumption-free way to perform data exploration is then to run TCA on the whole brain volume and over the entire sequence of slice-timed volumes. While this provides a screening of the data for detecting coarse artifacts, not all functionally induced spatio-temporal structure might be segregated in clusters under such a broad scope. Particularly, event-related responses, short in duration, and interspersed with regard to event type, are unlikely to be detected by unrestricted TCA.

TCA can be restricted in space and time. By applying TCA only in a partial volume of the brain one can focus on regions of interest, or disregard regions of

5. To further minimize interpolation errors, recent techniques such as the expansion with phase-invariant Fourier-sets as base functions, or the inclusion of temporal derivatives could be used, see (Henson et al., 1999; Josephs and Henson, 1999).

no interest. For uncovering functionally related structure more specifically, one can restrict TCA in the time dimension, i.e., inspect temporal segments of the data that have been selected informed by the experimental paradigm. We used temporal restriction for isolating effects from experimental conditions, in our case, from different event types. The selection of segments to be analyzed used the labeling of the measurement sequence as described in section 5.4.1.1 (Wichert et al., 2001a).

In the reminder of this section we will briefly describe two more technical topics that are essential components of the data analysis technique, the clustering method (section 5.4.2.2), and definitions of temporal and spatial similarity we used in the cluster analysis and to assess and compare the clustering results (paragraph 5.4.2.3). After TCA is carried out for each conditions independently, these similarity measures can be used to search for relations and differences in functional activity of different conditions.

5.4.2.2 Dynamical Clustering

As a number of previous studies have revealed, the standard clustering algorithm, k-means clustering is not the right choice for high-dimensional fMRI data sets (with hundreds of sampling points in time). The gradient-descent performed by the k-means algorithm is a local minimizer that for high-dimensional data sets frequently fails to find the global minimum, resulting in poor data fits. An indicator for the local minima problem of k-means clustering is a poor reproducability of the results. The results depend strongly on the initialization of the cluster centers. With random initialization[6] repeated runs of k-means clustering on the same data set can lead to quite different partition results.

Therefore, alternative algorithms have been proposed, like fuzzy clustering (Scarth et al., 1995; Moser et al., 1997; Baumgartner et al., 1997; Golay et al., 1998; Fadili et al., 2000), hierarchical clustering (Goutte et al., 1999; Filzmoser et al., 1999), and dynamical cluster analysis (DCA) (Baune et al., 1997, 1999). In DCA the number of clusters is not fixed like with k-means clustering; cluster centers are generated and anihilated during the data fitting process. A comparison between k-means CA and DCA has shown that the reproducability of DCA is much better (Baune et al., 1999). The problem with DCA is that it is computationally expensive. The advantage of dynamical cluster generation in DCA led us to a variant of k-means clustering with dynamical initialization phase for the choice of cluster seeds, very similar as the one proposed by Waldemark (1997). For a previously specified radius r the initialization phase generates a set of cluster seeds such that for every data point at least one of the seeds is closer than r. The initialization is completed after a single sweep through the data set where successively each data point is assigned as cluster seed if all previously assigned seeds are farther away

6. In random initialization, the default used for k-means clustering, k data points are picked at random as seeds for the cluster centers.

than r. We will refer to the combination of k-means clustering with the described dynamical initialization phase as K-MDI clustering. On the data of the working memory study both methods, K-MDI and DCA, achieved similar results, but the first was considerably faster. Therefore, in the following we will only report the results provided by K-MDI clustering.

5.4.2.3 Assessment and Comparison of Cluster Solutions

TCA requires one strong a priori assumption which is the measure of *temporal dissimilarity (td)* used for clustering. We used the Euclidean distance between two time courses. Another commonly used measure, that only compares signal shapes and entirely disregards absolute signal amplitudes, is based on the correlation coefficient.

After completion of TCA there is the problem how to inspect, assess and interpret the results in a systematic and fair manner. This requires quantitative description of the results, i.e., of the properties of centers and spatial patterns of the clusters. The *td* measure used during clustering can also be applied post hoc to the cluster centers for assessing differences in the temporal signal shape. In addition, we used the following definitions (Wichert et al., 2001a): The *signal change homogeneity (SCH)* which is defined as the ratio between peak-to-peak amplitude of the cluster center and mean standard deviation of signal amplitudes of the members of a cluster. The *temporal smoothness (TSM)* which is defined as the relative spectral power in the low frequency range of a time course. The *spatial contiguity (SC)* which is defined as the relative number of adjacent voxels in a voxel set (presence of "spatial clusters"). The *spatial similarity* ($0 \leq ss \leq 1$) between two patterns which is defined as the normalized overlap, i.e., the ratio between the numbers of voxels in the intersection and the union of the voxel sets corresponding to the clusters. For a pattern b and a pattern set A we call $a \in A$ *best match to* b, if a is the element with maximum spatial similarity to b. For two pattern sets A and B we call $a \in A, b \in B$ *best-matching pair*, if a is best match to B and b is best match to A.

For inspection and interpretation of clustering solutions we select clusters based on threshold criteria using the introduced measures. Functional activation can be assessed by focussing on clusters with high values in SCH and TSM. The first criterion selects clusters whose voxels homogeniously display signal changes of the cluster center. To assess homogeneity of time courses in fMRI activity maps Kendall's coefficient of concordance has been proposed (Baumgartner et al., 1999, 2000b). We prefer the SCH measure for cluster selection because it measures the homogeneity of signal changes which are essential for characterizing functional activity [7]. A threshold on TSM is used to reject clusters whose signal shape cannot be explained by influences mediated by the low-frequent HR.

7. For visual inspection of cluster homogeneity we display cluster centers with error bars reflecting standard deviation of signal amplitudes of the members.

As an optional selection criterion one can require high SC. This rules out clusters whose spatial patterns are scattered. Such rejection follows a common assumption that meaningful fMRI activity occupies a larger region than the volume of a single voxel (typically $1 \times 1 \times 3$ mm^3) and thus forms a *spatial cluster*[8]. This assumption also underlies spatial smoothing of the data, a preprocessing step, that is often applied before starting with data analysis (Xiong et al., 1995). Spatial filtering, for instance, with Gaussian kernels, enhances contiguous components in the signal by spatial averaging. In our technique spatial clustering is used as post hoc criterion to assess the result of temporal clustering[9].

The measures *ss* and *td* will be used for various comparisons between conditions. For instance, cluster results from two different conditions can be checked for spatial similarity. A spatially corresponding cluster pair (a best-matching pair) suggests that similar regions are recruited by both conditions. Differences in the centers of corresponding clusters indicate condition-specific changes of functional activation. Size differences of corresponding clusters, however, cannot be interpreted directly. Due to the global nature of the clustering process, the size of a cluster is influenced by push-away effects from other clusters[10].

5.5 The Working Memory Study

5.5.1 The Experimental Task and Data Aquisition

Five male and four female volunteers performed a delayed match-to-sample task in the fMRI scanner. Experiment blocks consisted of 42 task events with two different event types pseudo-randomly shuffled, one with low memory load (memory set with 1 letter) and one with high load (6 letters). Each event type occurred $r = 21$ times during one experimental block. An event started with the visual presentation of the memory set, a 2×3 array of the letters. For low load, a single letter to be memorized in the array was marked with a different color. The presentation lasted for 1s and 3.5s for low and high load, respectively. To reduce confounding effects in the following delay phase we adjusted memory set presentation time according to the number of items to be remembered (0.5s per item plus 0.5s) as usually done in behavioral studies (Richardson et al., 1996; Neath et al., 1999). After the delay

8. It is important to distinguish between spatial clusters, defined by voxel contiguities in spatial patterns, and the clusters extracted by clustering signal time courses as described in section 5.4.2.2. The latter are formed without any information about voxel contiguity. If the voxel pattern forms a spatial cluster this reflects the additional property that similar time courses are found in nearby voxels.
9. We apply spatial smoothing for preprocessing only before group averaging to account for imprecision of realignment and normalization and for interindividual differences in functional localization.
10. Neighboring clusters will compete for data points, resulting in a repulsive interaction.

period of 6s, a second visual stimulus was presented for 1.5s. It displayed a similiar 2×3 array containing one letter and five dummies. The subject had to press a yes/no button deciding whether or not the letter was in the memory set previously seen. Video goggles were used for visual presentation.

An experimental session consisted of two identical pseudorandom event blocks, as described above. For one male subject the experiment block was repeated five times. This data set was used in the single subject analysis. Data aquisition was performed on a 1.5 T Siemens Vision scanner. A full brain volume consisted of $k = 21$ slices and was sampled with TR = 1.9s.

5.5.2 Data Preprocessing and Analyses

We started data processing by motion correction using the realignment procedure in SPM99 (http://www.fil.ion.ucl.ac.uk/spm). In a second step for each subject the signals from repeated blocks were averaged. For assessing group effects the data of eight subjects (four male and four female) where spatially smoothed (with a Gaussian kernel of 8mm), normalized to the SPM Epi template, and averaged. Having performed the experiment with the DLS-fMRI technique described in section 5.4.1.1, there were two options for volume slice timing, DLS-re-sorting, or the usual phase-shift method. DLS-re-sorting was provided by a self-implemented program. For the purpose of comparison we also carried out the second option: The common signal interpolation slice timing was performed by the routine available in SPM99. Both, the phase-shifted and the DLS-re-sorted data set were analyzed in two different ways. We will use the following descriptors for different exprimental conditions: H denotes the high and L the low load condition in the working memory task. Different phases in the trials are denoted by s for memory set presentation, d for delay phase, and t for the test or probe phase. Thus, the set of different experimental conditions is $\mathcal{C} = \{Hs, Hd, Ht, Ls, Ld, Lt, P\}$, with P denoting pauses between events.

For inferential data analysis regression analysis in the GLM was carried out with SPM99. As regressors we used the box cars of the distinct experimental periods, stimulation, delay and target, convolved with a canonical HR function, a gamma function with a time-to-peak constant of 5s. The contrast functions we used in the GLM will be given in the result sections using a notation with the elements of $\mathcal{C}$. For instance, $Hd - Hs$ denotes the contrast function requiring a higher signal in the delay than in stimulus during the high load condition.

Cluster analysis was applied on different temporal segments of the data sets. For overall data exploration we applied TCA on $H\&L$, the data set containing the measurements during both load conditions. Such overall data exploration revealed for eight of the nine subjects scanner artifacts localized in the basal parts of the brain. For assessing group effects we disregarded the regions with distorted signal in the further analysis by spatial restriction of the cluster analysis. As described in section 5.4.2.1, we applied TCA separately on two data segments: L with a duration of 12s corresponding to the low load event; and H with a duration of 13.5s

corresponding to the high load event. The cluster analyses yielded 20 clusters for low load, and and 18 clusters for high load. In the following, these cluster sets will be labeled by $L1 - L20$, and $H1 - H18$, respectively.

5.5.3 Results for a Single Subject

This section compares the results from different data acquisition and analysis techniques for a single subject. The two most contrasting analysis approaches we employed were linear regression analysis (GLM) on the phase-shifted data set, as the most conventional, and K-MDI clustering on the DLS-data, as the most unconventional. If these extremes yielded agreeing results, we will not describe the results produced by intermediate approaches, such as TCA on phase-shifted data, or linear regression in the DLS data set. Overall data examination with TCA revealed no obvious artifacts so that the whole data set could be examined.

Our first question concerned visual functional activity. In the GLM we looked for voxels showing the same activation pattern in both load conditions, a higher signal intensity under visual stimulation, in the periods s and t, than without visual stimulation in the delay periods d. The binarized SPM_t map is shown in figure 5.2.

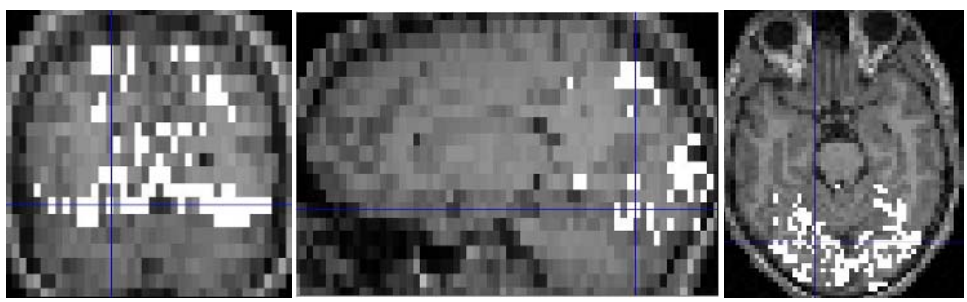


Figure 5.2 The SPM_t map of the contrast function $+Ls - Ld + Lt + Hs - Hd + Ht$ (corrected with $p < 0.05$) in the signal interpolation slice-timed data.

TCA was applied on the $H\&L$ DLS-data. For comparison with the results of standard GLM analysis we selected those clusters from the clustering result whose centers had highest temporal similarity with the used contrast function[11]. Note, however, that the selection of visual clusters was done post hoc, i.e., the formation of clusters was uninformed of any target function. The spatial maps of the two visual clusters are displayed in figure 5.3. The two clusters had very similar signal shape but different signal amplitude levels.

11. A boxcar function convolved with a canonical HR function.

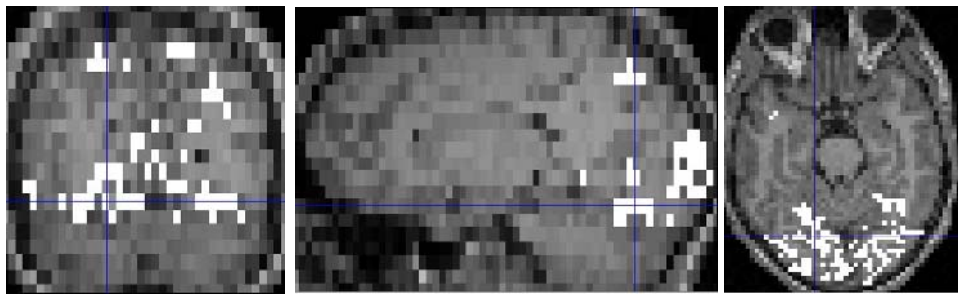


Figure 5.3 The clusters with highest temporal similarity to the contrast used in the SPM_t map displayed in figure 5.2.

Figures 5.2 and 5.3 show quite similar spatial maps for visual stimulation indicating that the time courses of visual stimulation were well preserved by both time-slicing operations. Further, this result suggested that visual activity is so dominant in the data statistics that it is lumped together even by a data exploration completely uninformed about the functional paradigm.

Our second question addressed delay activity, the focus of our study. We concentrated on the high load, using the contrast function $-Hs + Hd$, and masked the SPM_t map exclusively by two other contrast maps: stimulus deactivation against the baseline, i.e., $-Hs + P$, and second, higher delay activity during low load than during high load, i.e., $+Ld - Hd$. For the single subject this search was negative, neither in the slice-timed, nor in the DLS data set we were able to find significant delay activity with SPM or the exploratory analysis technique. Our negative result of finding delay-related functional activity in the single subject is in line with some previous fMRI studies of working memory fMRI, however, there are also studies reporting positive results.

Finally, we asked about functional activity related to both phases, delay period s and probe phase t. We applied a contrast including target and delay in both conditions against baseline $+D + T - P$. The obtained SPM_t map was exclusively masked by the contrast function for visual activation (cf. Figs. 5.3 and 5.4). The resulting map on the phase-shift slice-timed data is displayed in figure 5.4. Functional activity is scattered but shows higher spatial density in left parietal regions and bilateral prefrontal regions.

To assess the influence of the slice timing methods on such a smaller effect, we analyzed the DLS-re-sorted data set with the equivalent contrast function, see figure 5.5. Although the main spatial clusters agree, the maps differ in detail. There is less lateralization in the DLS-re-sorted data set than in the phase-shift slice-timed data set. This example shows that effects less salient than visual stimulation are influenced by the method of time-slicing.

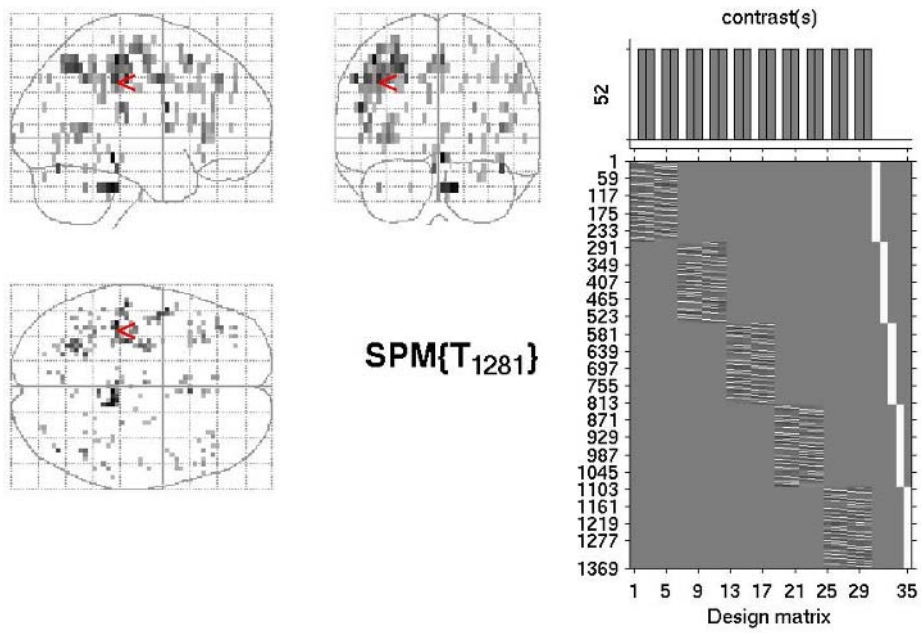


Figure 5.4 SPM_t map of the contrast function $+Ld + Lt + Hd + Ht - P$, exclusively masked by $+Ls - Ld + Lt + Hs - Hd + Ht$ (corrected. $p < 0.05$). Height threshold T = 5.12, extent threshold k=0. Time-slicing done with phase-shift method.

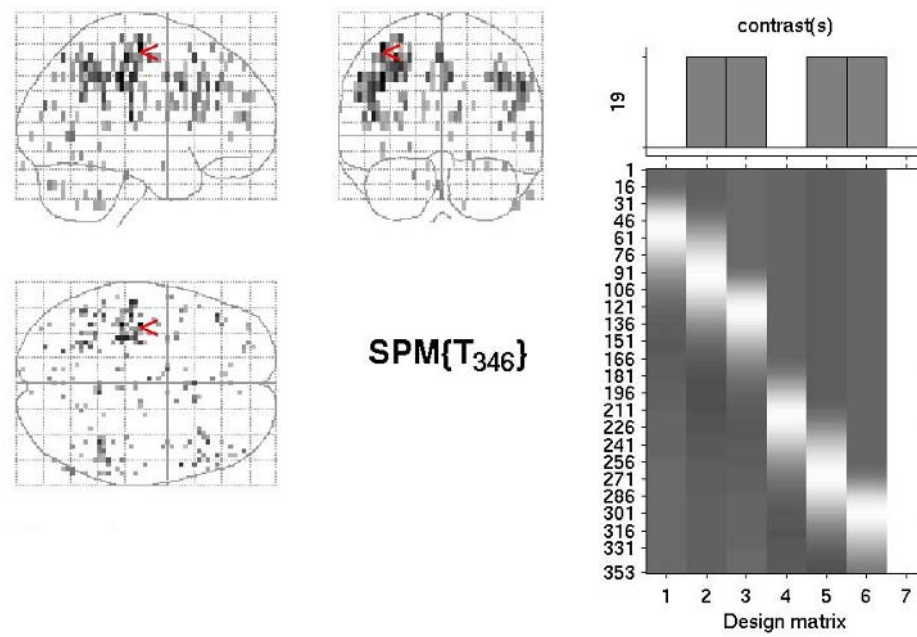


Figure 5.5 SPM_t map of the contrast function $+Ld + Lt + Hd + Ht - P$ exclusively masked by $+Ls - Ld + Lt + Hs - Hd + Ht$ (corrected, $p = 0.05$). Height threshold T=5.04, extent threshold k=0. DLS-re-sorted data.

5.5.4 Results of the Group Analysis

A group analysis can reveal effects similarly expressed in several group members, even if hidden by signal variability in individual data sets. The rationale of a group analysis is simply that group effects add up, while those signal components are averaged out that vary over the subjects.

Linear regression in the GLM was again used on the group data for cross-checking. We used the same combination of contrasts in the GLM as for the single subject analysis in section 5.5.3. Unlike in the single subject, where no functional activity could be detected, delay activation was found in the group analysis, see figure 5.6.

The cluster centers resulting from the TCA of the group data are displayed in figure 5.8. First we asked for spatial similarities between the high load clustering results and the SPM_t map of figure 5.6. The best and the second best matches to the SPM_t map were the clusters $H12$ and $H10$, reaching together a spatial similarity to the SPM_t map of $ss = 0.077$. Their spatial maps are displayed in figure 5.7.

Figures 5.6 and 5.7 reveal a qualitative agreement. Both spatial maps were dominantly located in the left superior parietal cortex (BA40), in regions at the midline, superior frontal gyrus (BA6) and in the left prefrontal cortex (BA9). Furthermore, the time courses of $H12$ and $H10$ were similiar to each other, and showed, in fact, delay activity; a pronounced peak in the second half of the delay period, see figure 5.8. Thus, the TCA found components with qualitative spatial

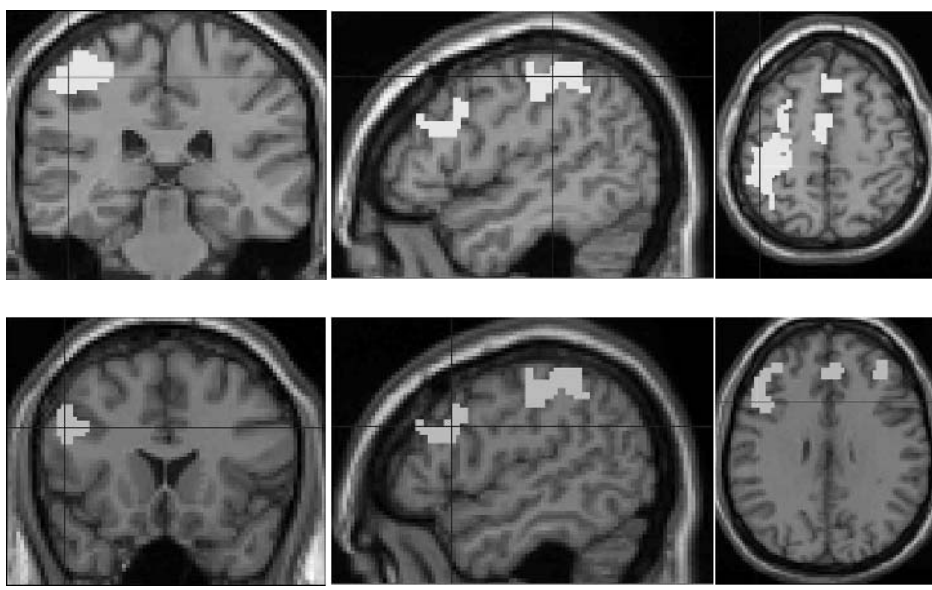


Figure 5.6 SPM_t map of the contrast $-Hs + Hd$, masked exclusively by $Hs - P$, and by $+Ld - Hd$ (corrected, $p = 0.05$). In the upper row the cursor position is in BA40, in the lower row in BA9. Phase-shift time-slicing used.

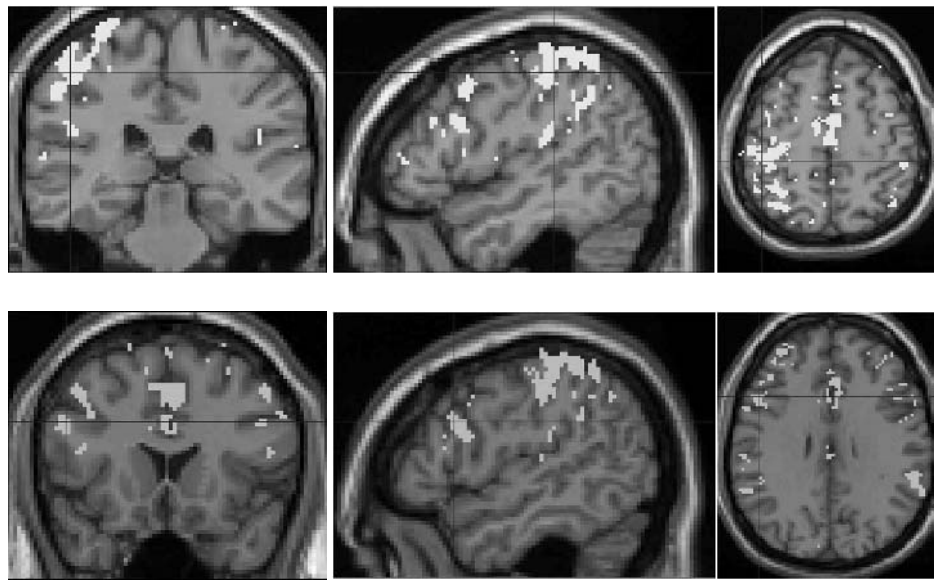


Figure 5.7 Spatial maps of $H10$ and $H12$, the clusters with highest spatial overlap with the SPM_t map of figure 5.6. Both cluster centers show a pronounced peak in the late delay period, see figure 5.8. In the upper row the cursor position is in BA40, in the lower row in BA9. DLS-re-sorting used for slice timing.

similarity to the SPM map. The detailed patterns of activity, however, deviated: the spatial similarity ss between the TCA clusters and the SPM_t stayed far below one, and at some regions they disagreed considerably. The spatial cluster located in right prefrontal cortex in the SPM map, for instance, had only some scattered voxels as counterparts in the TCA result. Differences must be caused by the different ways of data analysis and preprocessing as well. One should be aware, that there is no "gold standard" for analyzing these data. The mutual masking of contrasts used to generate the SPM_t map is just another way of exploratory data analysis.

The crucial question is now what the exploratory analysis technique based on cluster analysis can reveal about the expression of delay activity exceeding the scope of GLM-based data analysis. For the systematic assessment of the entire clustering results we proceeded as described in the sections 5.4.2.3 and 5.4.2.1. First, we selected clusters with high signal change homogeneity. The chosen selection threshold was $SCH \geq 2.6$, which was surpassed by 7 clusters for high, and 8 for low load condition. One low load cluster was excluded because of lacking temporal smoothness. The cluster centers of 11 of the selected clusters are displayed in figure 5.8. Among the selected high load clusters were the previously described "delay activity" clusters $H12$ and $H10$, but also $H15$, with quite similar activation time course, peaked in the late delay. There were more clusters displaying activation during the delay, however, with quite different time course: The signal courses of $H2$ and $H16$ show not only activation in the delay period but also in the phase of

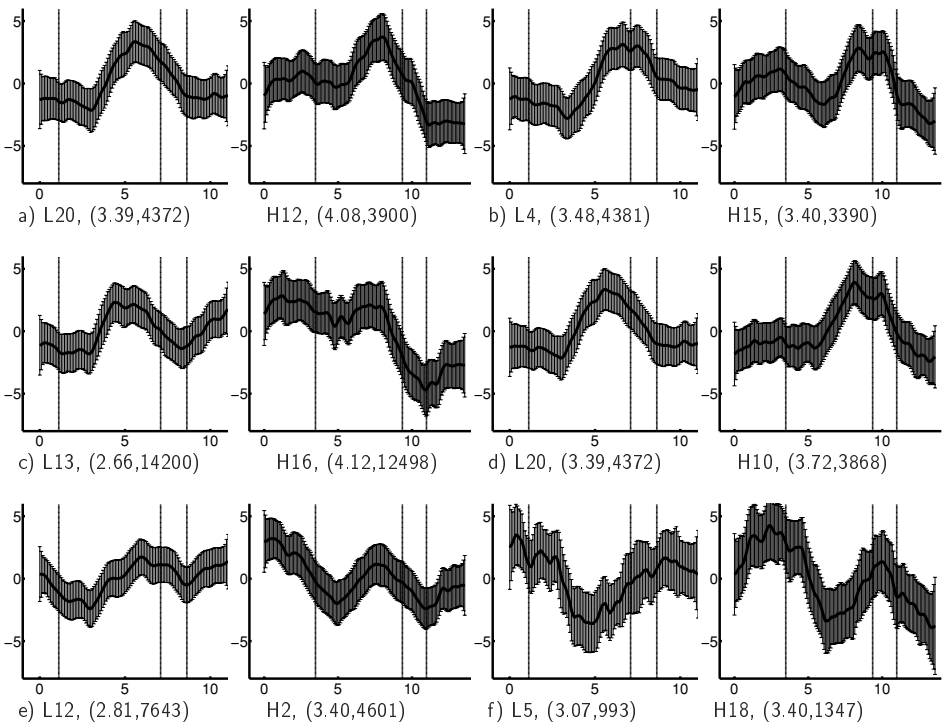


Figure 5.8 Cluster centers obtained by the analyses of high and low load trials. The displayed clusters satisfied our selection criteria based on the measures SCH and TCH (defined in section 5.4.2.3). High and low load clusters are paired with respect to high spatial similarity. Each pair a) to f) shows on the left the center for low load and on the right the center for high load. The horizontal axes display latency time with respect to stimulus onset (in seconds, $t = 0$ marks the canonical hemodynamic delay interval of 5s after stimulus onset). The curve onsets mark the begin of the phase s, the three bars indicate transitions between the phases s, d, t and P. The vertical axes display relative signal strength (in arbitrary units). Error bars represent the signal standard deviation within the cluster. Below each diagram one finds the name of the cluster, and in brackets, the SCH value and cluster size (in voxels). The pairs a) to f) are ordered with respect to decreasing spatial similarity (ranging from $ss = 0.156$ for a), to $ss = 0.127$ for f).

memory set presentation. The time course of $H2$ displayed an activity peak in the late delay period quite similar to the cluster centers of $H12$, $H10$ and $H15$, but there was a second activity peak in the preceeding period with visual stimulation. Of course, voxels with such temporal behavior remain undetected by the $d - s$ contrast used in the GLM-based data analysis. Figure 5.9 shows the spatial pattern of the cluster $H2$. The spatial pattern of cluster $H2$ was located in the upper left and right occipito-parietal cortex, (BA19, BA40).

Cluster $H16$ showed a signal time course that was quite unique among the clusters with increased activities in the delay. The signal was high during the visual stimulation and the activity persisted without interuption almost until to the end

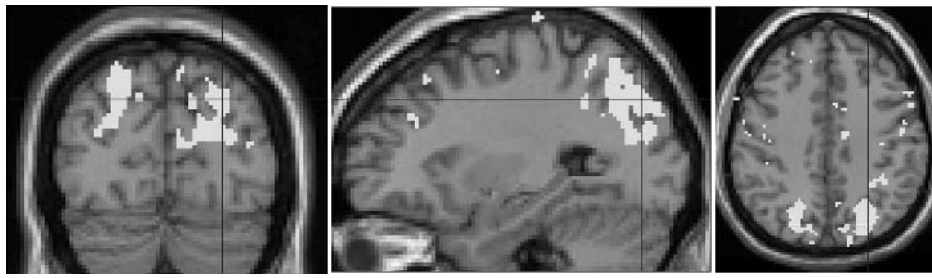


Figure 5.9 Map of cluster $H2$. It was activated in the stimulus and had a second peak in the delay. The cursor is set at the Talairach coordinates 27 -78.1 37.6 mm in BA19.

of the target period. The interpretation of this activation function is less clear since the prolonged activity could just be a confounding effect from the preceeding phase, but see discussion below.

So far, we have discussed the high load clusters showing pronounced effects in the delay period. To study load dependencies, we checked for spatial correspondences between high and low load clusters. We scored high/low load pairs with respect to spatial similarity. The spatial similarities ranged between $ss = 0.156$ and $ss = 0.0$. Interestingly, all 14 selected clusters belonged to the 10 high scored pairs, with spatial similarities higher than $ss = 0.098$. Eight of the 10 high scored pairs had a pairwise association, i.e., they were best-matching pairs, see definition in section 5.4.2.3. Figure 5.8 shows the six pairs with highest spatial similarity. With the exception of pair d), the clusters displayed in figure 5.8 were best-matching pairs. The pairs a) and d) associate the clusters $H12$ and $H10$ with the same low load cluster $L20$. Note that $H12$ and $H10$ were the both clusters with highest spatial similarity to the GLM map, see figure 5.7. The magnitudes of the spatial similarity within pairs already indicated a substantial overlap region[12] and a visually salient similarity.

Interestingly, in many cluster pairs, a), b), d), f), we found high similarity in signal shape and amplitude of the cluster centers. This indicated load-independent activation time courses in the overlap regions of these cluster pairs. The strongest load-dependency was observed in the overlap of pair c): While during low load the activity was peaked in the late delay ($L13$), the cluster $H16$, described above, showed persistent activity from stimulus to the end of the delay phase during high load. Since quite clearly $L13$ displayed delay-related activity, the region where $L13$ overlapped with $H16$ was likely to convey delay-related neuronal activity also during high load, even if this was unclear from the time course of $H16$ alone. The spatial pattern of the overlap region of cluster pair c) is shown in figure 5.10. It is spatially clustered in the precentral gyrus (BA6), bilateral near the midline (SMA), and left parietal cortex.

12. For instance, the overlap region of cluster pair c) is displayed in figure 5.10.

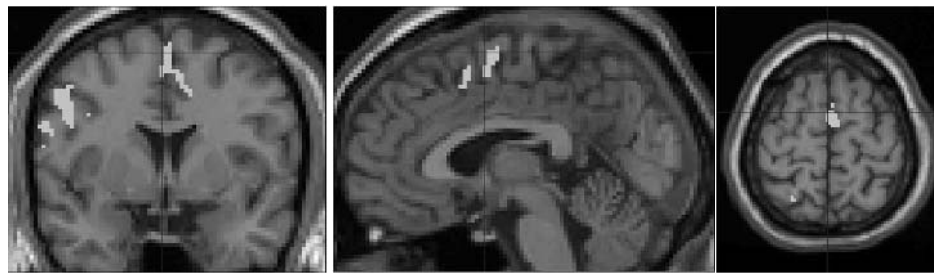


Figure 5.10 Voxels belonging to both clusters L13 and H16 of cluster pair c) showing the most extreme load-dependency in the activation. The cursor is set at the Talairach coordinates -2.2 1.4 61.0 mm in medial precentral gyrus (BA6).

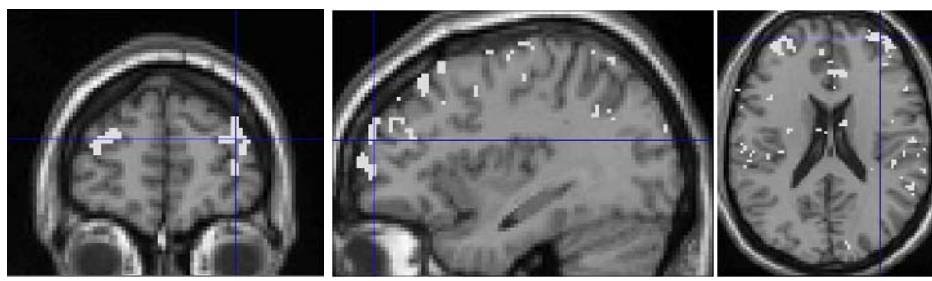


Figure 5.11 Load-dependent delay period specific effect. This map shows the voxels in the superset of high load clusters with delay peak (H10, H12, H15), but not in the superset of low load clusters with delay peak (L4, L20). The cursor is set at the Talairach coordinates 35.3 56.1 18.6 mm in the right BA10.

The spatial deviations within the cluster pairs cannot be interpreted directly, see section 5.4.2.1. For a rough estimation how the global region of exclusive delay activation changed with load we merged for each load all the clusters with the typical activity peak in the late delay/early target (the clusters in the pairs a, b, and d). The voxel ratio between high and low load was 1.27, indicating a slight load-dependent increase of the activated volume. We asked for the spatial distributions of load specific activation: The map of voxels with delay activation only in the high load, shown in figure 5.11, was clearly spatially clustered in prefrontal cortex and in gyrus cinguli anterior. The map of voxels with delay activation exclusively during the low load exhibited less pronounced spatial clusters located in premotor and posterior parietal regions (map not shown here).

5.6 Discussion

5.6.1 Relations to Other Exploratory Analysis Techniques for fMRI

A number of different exploratory analysis methods have been proposed for fMRI, e. g. cluster analysis, principal/independent component analysis, neural networks. These exploratory approaches have demonstrated that for block designs functional activity can be detected completely uninformed by the experimental paradigm, see references in section 5.4.2.

For event-related designs involving short events and different conditions, functional activation is unlikely to be found completely uninformed of the paradigm. Event-related effects are small compared to other signal influences and functional recruitment of different conditions might interfere with each other. Earlier studies applying explorative techniques on event-related experiments used previous knowledge about the task by introducing spatial constraints, i.e., restricting the analysis to regions of interests (Richter et al., 2000). The incentive of our technique is to explore the whole brain, by relying on temporal constraints dictated by experimental paradigm. In the case of the working memory study we applied TCA separately on data segments that corresponded to different load conditions.

Exploratory methods, though reducing raw data, still tend to produce data sets rich in structure (like the clusters obtained from temporal clustering as involved in our technique). Often, the a posteriori assessment of exploratory results is done more or less ad hoc and the selection of reported clusters does not rely on fair data-based criteria, but is biased by expectations about the results. We believe that a crucial component of exploratory data analysis is a systematic and fair a posteriori assessment of the reduced data sets. Therefore we introduced quantitative measures and similarity relations to select, evaluate and interprete the TCA clusters. We selected clusters with high signal change homogeneity SCH and high temporal smoothness TSM to account for functional activity[13]. However, once thresholds were fixed, all clusters were reported that fullfill the criterion.

The paradigm-informed application of TCA presented in this chapter actually increased the difficulty of a posteriori assessment. Instead of just one data partition we had to deal with different data partitions for each load condition. For a systematic assessment of load-dependent effects we scored the spatial similarity between cluster pairs. We found pairwise correspondences between most of the previously selected clusters. Thus, the assessment of load-dependencies mainly involved pairwise comparisons between clusters.

Various temporal clustering methods have turned out to perform well on fMRI data, see section 5.3. Biased by our previous experience, we chose a dynamical hard clustering algorithm (DCA) (Baune et al., 1999) and compared this approach with

13. In the displayed mean time courses we visualize by error bars the cluster homogeneity proposed earlier (Baumgartner et al., 1999).

faster clustering algorithms. Previously it had been shown that k-means cluster analysis with random initialization of the cluster seeds has lower robustness than DCA (Baune et al., 1999). We found that k-means clustering, if extended by an dynamical initialization phase (Waldemark, 1997), can achieve similar results and similar robustness as DCA. The dynamical initialization proved to be effective in preventing the gradient-descent to terminate in local minima.

Volume slice timing is an inevitable step of data preprocessing for exploratory data analysis based on temporal similarity. In common slice timing techniques (Aguirre et al., 1998) the signal interpolation is a potential source of signal distortion. Since the detection of functional activation during a working memory task with short delay is hard (low S/N), we wanted to eliminate as much as possible sources of noise in data preprocessing. We proposed dense latency sampling (DLS) that not only enhances time resolution by oversampling (Josephs et al., 1997; Miezin et al., 2000), but provided slice timing without introducing any signal interpolation errors. It should be noted that the re-sorting process of the DLS-technique destroys the auto-correlation structure of signal components that are unrelated to the events. However, this does not have a strong impact on the detection of event-related activity.

5.6.2 Probing Delay Related Activity: Methodological Issues

The slow HR characteristic is the limiting factor for the temporal resolution of regional cerebral blood-flow-based techniques including fMRI. A serious problem these techniques have with event-related data acquisition is to disentangle functional activity from different events following on each other in fast succession. In the case of event-related working memory studies, delay related activity might be confounded by the HR resulting from processes during the preceding presentation period. A way to exclude this confound is to prolong the delay period – e.g. to 12s – and to consider only delay activity occuring not earlier than 4-5s after the onset of the delay period, e.g. (Zarahn et al., 1997; Cohen et al., 1997; Postle et al., 1999; Rypma and D'Esposito, 1999; D'Esposito et al., 2000). However, this approach has major drawbacks, such as the qualitative changes of working memory with variation of the delay period, and the fact that experimental sessions become quite long.

In this study we examined a delayed response task with a delay period of 6s. The chosen delay length was still in the time domain of working memory most intensely studied by other experimental methods. Though the delay duration was on the short end of those examined by other neuroimaging studies, we were confident to detect delay-related activity for a number of reasons. Visual experiments had indicated almost linear summation for the hemodynamic responses of different events (Boynton et al., 1996; Buckner, 1998), and consecutive events have been successfully resolved, even if their time delay was only about $2 - 3$s (Kim et al., 1997). Burock et al. (1998) demonstrated that at fast stimulus presentation rates the hemodynamic responses could be estimated quite well. Moreover, confounding due to the HR is only a problem if successive events activate the same regions. Small

latencies (of the order of several hundreds of milliseconds) between different regions had been previously detected, for instance in regions involved in voluntary hand movement (Wildgruber et al., 1997; Baune et al., 1999). To provide the optimum power for detection of delay-specific activity, we tried to optimize experimental paradigm, design, and the exploratory technique: i) In order to restrict in both load conditions encoding processes to the period of memory set presentation, we adjusted the presentation time according to the number of items to be remembered, see section 5.5. ii) We eliminated possible distortions from signal interpolation by applying the DLS method. iii) In the GLM analysis we checked for delay activity against the activity level during stimulus presentation, and not against baseline. iv) The exploratory technique was unbiased by a priori hypotheses about the signal shape of functional activation. Such a bias reduces sensitivity in case of mismatch between hypothesis and actual activation.

5.6.3 Results on Working Memory

We examined a delayed response task with 6s delay duration and two different load conditions (1 versus 6 letters). In a single subject we found functional activation related to the phase including delay period and probe phase in parietal and prefrontal regions. However, none of the techniques found purely delay-related activation. The situation was different in the data averaged over 8 subjects. Here clear delay-specific activity was found, with the conventional slice timing and analysis technique (figure 5.6) and as well with the exploratory analysis technique (figure 5.7). Both analysis techniques located the delay activity accordingly in typical working memory regions, inferior parietal left, superior frontal left, superior medial (BA6) and bilateral prefrontal (BA9).

The exploratory technique yielded results going beyond those obtained by an inferential technique, mainly because the former can characterize functional activation unbiased by expected signal shapes (regressors) or locations (anatomically defined regions of interest[14]). Most delay-related activity found by the exploratory data analysis had a particular transient signal shape, a peak in the second half of the delay period (figure 5.8 a,b,d,L13). In some clusters the peak width was somewhat wider and involved some of the target period too (figure 5.8 b,H10). Since all these time courses involved no high activity in the preceeding visual stimulus period, confounding effects were no problem in indentifying these activity peaks as delay-related. In pre- and sensomotoric regions delay peaks were found during both load conditions (overlaps of the cluster pairs a, b and d in figure 5.8). In other regions delay peaks occurred load specifically: for high load in bilateral spatial clusters in PFC, and in anterior cingulate (figure 5.11). The latter regions seemed to be particularly involved in working memory related processes during the delay pe-

14. Jhi and McCarthy (2000) studied signal time courses during delayed tasks averaged over anatomically defined regions of interest.

riod. Since the activation occured in the later part of the delay period, they might participate in rehearsal as well as in preparation of the stored information to be used in desicion making. The left lateralization of delay-related activity (seen in Figs. 5.6 and 5.7) is in accordance to a number of studies on the processing of verbal material, see for instance (Awh et al., 1996; Gabrieli et al., 1998; Smith et al., 1997; Smith and Jonides, 1999).

Other clusters showed activity during the delay period as well, but the activation also included the preceeding period of memory set presentation (figure 5.8 $H2$,$H16$). The functional interpretation of these clusters relied on their detailed time courses or on spatial relationships between clusters of the different load conditions. One cluster center ($H2$) exhibited two peaks clearly separated in time, one during stimulus presentation and one during delay. This suggested that the delay activity is no confound from the preceeding phase, but a true delay-related component. The cluster was localized in bilateral superior parietal regions near midline. We conclude that superior parietal regions are not only involved in perception and encoding, but also in delay-locked processes of working memory, such as rehearsal. Another cluster ($H16$) allowed no functional assignment based on its signal shape, the activity was high during visual stimulation and persisted continously almost over the complete delay period. But for this cluster the best-matching cluster from the analysis of low load events ($L13$) showed peaked delay activity (figure 5.11 c). Thus, voxels in the overlap of $H16$ and $L13$ are likely to be involved in delay-related processes as well. These voxels were located in BA6, precentral and left lateral (figure 5.10). In BA6 the load dependency was largest during the period of visual stimulation. Thus, BA6 might participate in working memory maintainance, but its most important role is more likely to be in encoding.

In this paragraph we summarized results obtained with the explorative technique and how they can be interpreted in terms of process specific involvements of cortical regions. A more complete description and discussion of the results in the context of process specific theories of working memory (Baddeley, 1996), as well as the relation to other working memory studies, will be subject of a forthcoming paper.

5.6.4 Conclusions

We have described a technique of exploratory data analysis for ER fMRI experiments. The technique includes two components. The first component is a new oversampling and data sorting scheme (DLS: dense latency sampling. The second component is a paradigm-informed application of temporal cluster analysis to event-related data combined with a systematic evaluation of the clustering results. A dynamical variant of k-means analysis (K-MDI CA) permitted rapid and reproducible cluster analysis of the data.

The technique was used for a study of delayed response working memory. As a reference, we also used standard techniques to evaluate the same data set. At a macroscopic level, both methods gave a similar view of patterns of delay-related approaches gave similar results about. Thus important features of the results can

be extracted by methods that differ in the type of underlying assumptions—functional specialization versus functional integration. Furthermore, the exploratory technique yielded results that the standard technique could not provide. The exploratory technique was able to provide a global view of the spatio-temporal structure associated with each different type of event. Thus, it was possible to assess involvement of disparate regions in different process of working memory.

The time course of the delay-related activity that we found differed from that measured with single cell recordings. The delay-related activity we measured reached a peak over time while the firing rates of individual prefrontal cells is persistent, a pattern that is interpreted as memory maintainance. The difference between the results might come from the fact that fMRI signals reflect activity of functionally diverse populations of cells that are involved in several processes. Such processes could include rehearsal that sets in later in the delay and computations that prepare the decision process.

In additon, our results suggest that superior parietal visual areas might be not only involved in perception or encoding, but also in rehearsal or decision making. This finding is especially encouraging because the involvement of sensory areas in decision processes has recently been found in electrophysiological studies of working memory, see (Brody, 2002).

We hope to have been able to convince the reader that using exploratory analysis for event-related fMRI adds an interesting alternative to existing techniques. Further we advocated that paradigm-informed application of TCA and systematic assessment of clustering results allows exploratory data analysis without restriction to regions of interest as by Richter et al. (2000). Finally, we proposed a solution of the slice timing problem (a problem that must be solved before any type of exploratory data analyis can be done) that avoids artifacts introduced by the standard method.

We believe it is important to add a final remark about one caveat of multivariate analysis of fMRI at high temporal resolution. Current methods rely on the assumption that synchrony of the hemodynamic response in different voxels corresponds to synchrony of the underlying neuronal activity. This assumption becomes more and more questionable as temporal resolution is increased. Indeed, there is some evidence that the delay of the hemodynamic response can vary with location (Aguirre et al., 1998) and with type of stimulation (Friston et al., 1998). Thus, further advances in fMRI will depend on a better understanding of the relationship between the hemodynamic response and patterns of neural activity.

Acknowledgments

This work was supported by a Landesforschungsschwerpunktprogramm grant from the state of Baden-Württemberg and by Wilhelm-Schweizer-Zinnfiguren in Diessen.

References

G. K. Aguirre, E. Zarahn, and M. D'Esposito. The variability of human BOLD hemodynamic responses. *NeuroImage*, 8(4):360–369, 1998.

E. Awh, J. J. Jonides, E. E. Smith, E. H. Schumacher, R. A. Koeppe, and S. Katz. Dissociation of storage and rehearsal in verbal working memory: evidence from positron emmission tomography. *Psychological Sciences*, 7:25–31, 1996.

A. Baddeley. *Working Memory*. Oxford University Press, Oxford, England, 1986.

A. Baddeley. The fractionation of working memory. *Proceedings of the National Academy of Sciences USA*, 93:13468–13472, 1996.

P. A. Bandettini, A. Jesmanowicz, E. C. Wong, and J. S. Hyde. Processing strategies for time-course data sets in functional MRI of the human brain. *Magnetic Resonance in Medicine*, 30:161–173, 1993.

R. Baumgartner, L. Ryner, W. Richter, R. Summers, M. Jarmasz, and R. Somorjai. Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis. *Magnetic Resononance Imaging*, 18(1):89 – 94, 2000a.

R. Baumgartner, G. Scarth, C. Teichtmeister, R. Samorjai, and E. Moser. Fuzzy clustering of gradient-echo functional MRI in human visual cortex. part I: Reproducibility. *Journal of Magnetic Resonance Imaging*, 7:1094 – 1101, 1997.

R. Baumgartner, R. Somorjai, and W. Richter. Assessment of cluster homogeneity in fMRI data using Kendall's coefficient of concordance. *Journal of Magnetic Resononance Imaging*, 17(2):1525 – 1532, 1999.

R. Baumgartner, R. Somorjai, R. Summers, W. Richter, and L. Ryner. Correlator beware: Correlation has limited selectivity for fMRI data analysis. *NeuroImage*, 12(2):240–243, 2000b.

A. Baune, M. Erb, F. T. Sommer, D. Wildgruber, U. Klose, and W. Grodd. Evaluation of fast fMRI measurements in motorcortex: Comparing a new cluster analysis with z-mapping. *Exp. Brain Res.*, 117:50, 1997.

A. Baune, F. T. Sommer, M. Erb, D. Wildgruber, B. Kardatzky, and W. Grodd. Dynamical cluster analysis of cortical fMRI activation. *NeuroImage*, 6(5):477 – 489, 1999.

G. M. Boynton, S. A. Engel, G. H. Glover, and D. J. Heeger. Linear systems analysis of functional magnetic resonance imaging in human. *Journal of Neuroscience*, 16:4207 – 4221, 1996.

C. Brody. Dynamics of sensory area during sensory discrimination and decision task. Talk presented at *NICE 2002 workshop*, Les Houches, 2002.

R. L. Buckner. Event-related fmri and the hemodynamic response. *Human Brain Mapping*, 6(5-6):373–377, 1998.

M. A. Burock, R. L. Buckner, M. G. Wolfdorff, B. R. Rosen, and A. M. Dale. Randomized event-related experimental desighs allow for extremely rapid presentation rates using functional MRI. *Neuroreport*, 9:3735–3739, 1998.

M. A. Burock and A. M. Dale. Estimation and detection of event-related fmri signals with temporally correlated noise: A statistically efficient and unbiased approach. *Human Brain Mapping*, 11(4):249–260, 2000.

R. Cabeza and L. Nyberg. Imaging cognition II: an empirical overview of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, 12(1):1–47, 2000.

J. D. Cohen, W. M. Perlstein, T. S. Braver, L. E. Nystroem, D. C. Noll, J. Jonides, and E. E. Smith. Temporal dynamics of brain activation during a working memory task. *Nature*, 386:604–608, 1997.

A. M. Dale. Optimal experimental design for event-related fMRI. *Human Brain Mapping*, 8(2-3):109–114, 1999.

A. M. Dale and R. L. Buckner. Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, 5(5):329–340, 1997.

M. D'Esposito, D. Ballard, E. Zarahn, and G. K. Aguirre. The role of prefrontal cortex in sensory memory and motor preparation: An event-related fMRI study. *NeuroImage*, 11(5):400–408, 2000.

M. D'Esposito, J. A. Detre, D. C. Alsop, R. K. Shin, S. Atlas, and M. Grossman. The neural basis of the central executice system of working memory. *Nature*, 378: 279–281, 1995.

M. D'Esposito, B. R. Postle, J. Jonides, and E. E. Smith. The neural substrate and temporal dynamics of interference effects in working memory as revealed by event-related functional MRI. *Proceedings of the National Academy of Sciences USA*, 96:7514–7519, 1999.

M.J. Fadili, S. Ruan amd D. Bloyet, and B. Mazoyer. A multistep unsupervised fuzzy clustering analysis of fmri time series. *Human Brain Mapping*, 10(4):160 – 178, 2000.

P. Filzmoser, R. Baumgartner, and E. Moser. A hierarchical clustering method for analyzing functional MR images. *Magnetic Resonance Imaging*, 17(6):817 – 826, 1999.

K. J. Friston, P. Fletcher, O. Josephs, A. P. Holmes, K. J. Worsley, M. D. Rugg, and R. Turner. Event-related fmri: characterizing differential responses. *Neuroimage*, 7(1):30–40, 1998.

John D. E. Gabrieli, Russell A. Poldrack, and John E. Desmond. The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences, U.S.A.*, 95:906–913, 1998.

X. Golay, S. Kollias, G. Stoll, D. Meier, A. Valavanis, and P. Boesiger. A new correlation-based fuzzy logic clustering algorithm for fmri. *Magn Reson Med*, 40: 249–260, 1998.

P. Goldman-Rakic. The prefrontal landscape: Implications of functional architecture for understanding of human mentation and the central executive. *Philosophical Transactions of the Royal Society B*, 351:1445–1453, 1996.

P. Goldman-Rakic. Localization of function all over again. *NeuroImage*, 11(5): 451–457, 2000.

C. Goutte, P. Toft, E. Rostrup, F. A. Nielsen, and L. K. Hansen. On clustering fMRI time series. *NeuroImage*, 9(3):298–310, 1999.

L. K. Hansen, J. Larsen, F. A. Nielsen, S. C. Strother, E. Rostrup, R. Savoy, N. Lange, J. Sidtis, C. Svarer, and O. B. Paulson. Generalizable patterns in neuroimaging: how many principal components? *NeuroImage*, 9(5):534–544, 1999.

R. N. A. Henson, C. Büchel, O. Josephs, and K. Friston. The slice-timing problem in event-related fMRI. *NeuroImage*, 9:125, 1999.

P. Jezzard and S. Clare. Sources of distortion in functional MRI data. *Human Brain Mapping*, 8(2-3):80–85, 1999.

A. P. Jhi and G. McCarthy. The influence of memory load upon delay-interval activity in a working-memory tast: an event related functional MRI study. *Journal of Neuroscience*, 12(Supplement 2):90–105, 2000.

J. Jonides, E. E. Smith, R. A. Koeppe, E. Awh, S. Minoshima, and M. A. Mintun. Spatial working memory in humans as revealed by PET. *Nature*, 363:623–625, 1993.

O. Josephs and N. A. Henson. Event-related funcitonal magnetic resonance imaging: modelling, inference and optimization. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 354:1215–1228, 1999.

O. Josephs, R. Turner, and K. Friston. Event-related fMRI. *Human Brain Mapping*, 5:243–248, 1997.

S.-G. Kim, W. Richter, and K. Ugurbil. Limitations of temporal resolution in functional MRI. *Magnetic Resonance in Medicine*, 37(4):631–636, 1997.

S. H. Lai and M. Fang. A novel local pca-based method for detecting activation signals in fmri. *Magnetic Resonance Imaging*, 17(6):827–836, 1999.

N. Lange. Statistical approaches to human brain mapping by functional magnetic resonance imaging. *Statistics in Medicine*, 15(4):389–428, 1996.

N. Lange. Empirical and substantive models, the Bayesian paradigm, and meta-analysis in functional brain imaging. *Human Brain Mapping*, 5(4):259–263, 1997.

N. Lange, S. C. Strother, J. R. Anderson, F. A. Nielsen, A. P. Holmes, T. Kolenda, R. Savoy, and L. K. Hansen. Plurality and resemblance in fMRI data analysis. *NeuroImage*, 10(3):282–303, 1999.

G. MCarthy, M. Luby, J. Gore, and P. Goldman-Rakic. Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI. *Journal of Neurophysiology*, 77:1630–1634, 1997.

M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. Kindermann, A. J. Bell, and T. J. Sejnowski. Analysis of fMRI by blind separation into independent spatial components. *Human Brain Mapping*, 6:160–188, 1998.

M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski. Functionally independent components of early event-related potentials in a visual spacial attention task. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 354:1135–1144, 1999.

M. J. McKeown and T. J. Sejnowski. Independent component analysis of fMRI data: Examining the assumptions. *Human Brain Mapping*, 6:368–372, 1998.

F. M. Miezin, L. Maccotta, J. M. Ollinger, S. E. Petersen, and R. L. Buckner. Characterizing the hemodynamic response: Effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImage*, 11(6):735–759, 2000.

P. P. Mitra, S. Ogawa, X. Hu, and K. Ugurbil. The nature of spatiotemporal changes in cerebral hemodynamics as manifested in functional magnetic resonance imaging. *Magn Reson Med*, 37:511–518, 1997.

E. Moser, M. Diemling, and R. Baumgartner. Fuzzy clustering of gradient-echo functional MRI in human visual cortex. part ii: Quantification. *Journal of Magnetic Resonance Imaging*, 7:1102–1108, 1997.

I Neath, G. D.A. Brown, M. Poirier, and C. Fortin. Short-term/working memory: An overview. *International Journal of Psychology*, 34(5/6):273 – 275, 1999.

A. M. Owen, A. C. Evans, and M. Petrides. Evidence for a two-pstage model of spatial working memory within the lateral frontal cortex: a positron emission tomography study. *Cerebral Cortex*, 6:31–38, 1996.

K. M. Petersson, T. E. Nichols, J. B. Poline, and A. P. Holmes. Statistical limitations in functional neuroimaging. i. non-inferential methods and statistical models. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 354:1239–1260, 1999a.

K. M. Petersson, T. E. Nichols, J. B. Poline, and A. P. Holmes. Statistical limitations in functional neuroimaging. ii. signal detection and statistical inference. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 354:1261–1281, 1999b.

B. R. Postle, , and M. D'Esposito. Evaluating models of the topographical organization of working memory function in frontal cortex with event-related fMRI. *Psychobiology*, 28:132–145, 2000a.

B. R. Postle, J. S. Berger, and M. D'Esposito. Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. *Proceedings of the National Academy of Sciences, U.S.A.*, 96:12959–12964, 1999.

B. R. Postle, E. Zarahn, and M. D'Esposito. Using event-related fmri to asses delay-period activity during performance of spatial and nonspatial working memory tasks. *Brain Research Protocolls*, 1:57–66, 2000b.

G. Rainer, W. F. Asaad, and E. K. Miller. Memory fields of neurons in the primate prefrontal cortex. *Proceedings of the National Academy of Sciences USA*, 95: 15008–15013, 1998.

J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, and T. Zacks. *Working Memory and Human Cognition*. Oxford University Press, Oxford, England, 1996.

W. Richter, R. Somorjai, R. Summers, M. Jarmasz, R. S. Menon, J. S. Gati, A. P. Georgopoulos, C. Tegeler, K. Ugurbil, and S.-G. Kim. Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *Journal of Cognitive Neuroscience*, 12(2):310–320, 2000.

B. R. Rosen, Randy L. Buckner, and Anders M. Dale. Behind the scenes of functional brain imaging: A historical and physiological perspective. *Proceedings of the National Academy of Sciences, U.S.A.*, 95:773–780, 1998.

J. B. Rowe, I. Toni, O. Josephs, R. S. J. Frackowiak, and R. E. Passingham. The prefronatal cortex: response selecthion or maintainance. *Science*, 288:1656–1660, 2000.

B. Rypma and M. D'Esposito. The roles of prefrontal brain regions in components of working memory: effects of memory load and individual differences. *Proceedings of the National Academy of Sciences, U.S.A.*, 96(11):6558–6563, 1999.

G. Scarth, M. McIntire, B. Wouk, and R. L Samorjai. Detecxtion of novelty in functional images using fuzzy clustering. *Proceedings of the Society of Magnetic Resonance*, 3rd Sci. Meeting:238, 1995.

T. Schanze. Sinc interpolation of discrete periodic signals. *IEEE Transactions on Signal Processing*, 43(6):1502–1503, 1995.

E. E. Smith and J. J. Jonides. Storage and execution in the frontal lobes. *Science*, 283:1657–1661, 1999.

E. E. Smith, J. J. Jonides, C. Marshuetz, and R. A. Koeppe. Components of verbal working memory: evidence from neuroimaging. *Proceedings of the National Academy of Sciences USA*, 95:875–882, 1997.

J. Waldemark. An automated procedure for cluster analysis of multivariate satellite data. *International Journal of Neural Systems*, 8(1):3–15, 1997.

A. Wichert, A. Baune, G. Grön, H. Walter, and F. T. Sommer. Interpretation of event-related fMRI using cluster analysis. In V. Kurkova, N. C. Steele, R. Neruda, and M. Karny, editors, *Artificial Neural Nets and Genetic Algorithms*, 446–448. Springer, Wien, 2001a.

A. Wichert, H. A. Kestler, H. Walter, G. Grön, A. Baune, J. Grothe, A. Wunderlich, and F. T. Sommer. Explorative detection of delay activity during a working memory task. In G. M. Papadourakis, editor, *Proc. 4th Int. Conference on Neural Networks and Expert Systems in Medizine and Healthcare*, 266–271. Technological Educational Institute of Crete, Heraklion, 2001b.

D. Wildgruber, M. Erb, U. Klose, and W. Grodd. Sequential activation of suplementary motor area and primary motor cortex during self-paced finger movement in human evaluated by functional neural MRI. *Neuroscience Letters*, 227:1–4, 1997.

J. Xiong, J.-H. Gao, J.L. Lancaster, and P. T. Fox. Clustered pixel analysis for functional MRI activation studies of the human brain. *Human Brain Mapping*, 3:287–301, 1995.

E. Zarahn, G. K. Aguirre, and M. D'Esposito. Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage*, 5(2):179–197, 1997.

6 Functional Magnetic Resonance Imaging Adaptation: A Technique for Studying the Properties of Neuronal Networks

Andreas S. Tolias
Max-Planck-Institute for Biological Cybernetics
72076 Tuebingen, Germany

Zoe Kourtzi
Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139

Max-Planck-Institute for Biological Cybernetics
72076 Tuebingen, Germany

Nikos K. Logothetis
Max-Planck-Institute for Biological Cybernetics
72076 Tuebingen, Germany

Functional magnetic resonance imaging can be used to study the networks of neurons that underline different behaviors. The blood oxygenation level-dependent signal though, measures the activity averaged across heterogeneous population of neurons with different response characteristics. It is therefore often impossible to infer the properties of the underlying imaged neural populations by simply examining the fMRI signal. Here, we describe the use of an adaptation paradigm to study the properties of neuronal populations beyond the spatial resolution of fMRI.

6.1 Introduction

A fundamental goal of research in systems neuroscience is to understand the neuronal mechanisms that underlie behavior, both at the level of single neurons

and of neuronal ensembles. Substantial progress has been made in characterizing the response properties of single neurons involved in sensory, motor and cognitive functions (Barlow, 1972). In contrast, little is known about the collective properties of contiguous or distributed networks of neurons that underlie brain mechanisms. Imaging techniques such as functional magnetic resonance imaging (fMRI), offer global coverage and could therefore be used to investigate brain mechanisms at the level of distributed networks of neurons. The fMRI signals are useful for studying the global organization of brain circuits, but are far removed from measuring the spike trains of individual neurons. In fact, recent studies have shown that the blood oxygenation level-dependent (BOLD) signal is correlated more with local field potentials and less with the average spiking activity of neurons (Logothetis et al., 2001). However, spikes are thought to be the language of the brain and the basic unit of neural computation. Therefore, the ability to combine fMRI with single cell recordings is a promising approach for systems neuroscience as the two techniques are complementary, providing information on different spatiotemporal scales. In addition fMRI can be used to localize putative brain circuits important for a particular experiment of interest so that these sites can be subsequently studied in depth with electrophysiology. Another limitation of standard imaging methods is its low spatial resolution. Namely, the measured signal represents the activity averaged across heterogeneous populations of neurons with different response characteristics. In most cases, it is therefore impossible to infer the properties of the underlying imaged neural populations. Recent studies have suggested that it may be possible to image brain activity organized in columns where nearby neurons have similar properties (Kim et al., 2000a; Kim et al., 2000b; see also Logothetis, 2000 for a critical evaluation). Even if this were the case, neuronal selectivity is not always organized in large columnar structures. Thus, given the functional architecture of the brain the conventional use of fMRI is rather restricted.

6.2 Adaptation Paradigm

Here we describe a novel adaptation paradigm that has been recently employed to study the properties of neuronal populations beyond the spatial resolution of fMRI. In this paradigm, sufficiently prolonged or repeated presentation of a stimulus (adapting stimulus) results in decreased fMRI responses. Once the underlying neuronal activity has been adapted, a test stimulus is presented which is either identical to the adapting stimulus or differs from it in specific attributes, (e.g. orientation, direction of motion etc). Stronger fMRI responses to the modified test stimulus rather than to the identical adapting test stimulus, indicate that neuronal populations within the imaged voxels are involved in processing information about the stimulus attribute along which the stimulus changed. To demonstrate more clearly the reasoning behind the use of the adaptation paradigm we propose a thought experiment as described in figure 6.1.

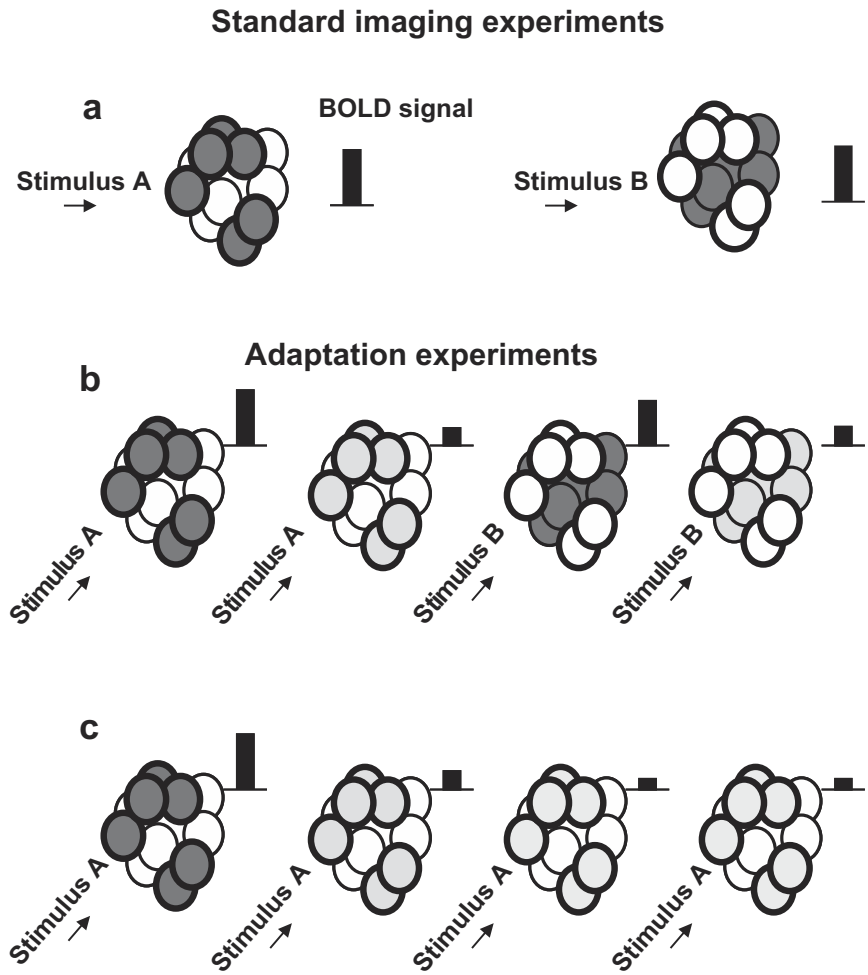


Figure 6.1 Thought experiment illustrating the use of an adaptation paradigm. (a) Conventional imaging experiment. During a conventional sensory fMRI experiment responses under two stimulus conditions A and B are compared to each other. Since the voxel of interest is composed of a heterogeneous population of neurons, i.e. not all neurons have the same stimulus selectivity, the strength of the BOLD signal will be the same under these two conditions. Therefore, this standard imaging experiment will fail to characterize the properties of these neurons. In this case the black and gray outlined neurons prefer stimuli A and B respectively. (b) Adaptation experiment. In this case stimulus A is shown for a prolonged time (or repeated). This results in adaptation of the BOLD signal. After presentation of stimulus B the signal shows a rebound, a sign of release from adaptation. This is due to the activation of the gray neurons for which stimulus B is their preferred stimulus. (c) In the control case of the adaptation experiment stimulus A is shown continuously instead of changing to stimulus B.

The goal of the thought experiment is to identify whether a particular voxel processes information about the direction of motion of a stimulus. If one were to carry out a conventional imaging experiment and compare the activity of this voxel under two stimulus conditions e.g. stimulus movement in two different directions, then no difference would be observed in the BOLD response signal (figure 6.1a). This would be due to the fact that this voxel is composed of a heterogeneous population of neurons; that is the neurons within this voxel do not have the same stimulus selectivity to motion direction. Therefore from such a standard imaging experiment, one cannot deduce the properties of the underlying neuronal population.

Now consider the same experimental question studied with an adaptation paradigm (figure 6.1b and 6.1c). After repeated presentation of a stimulus moving in the same direction of motion (e.g. leftward), the neuronal populations selective for this stimulus will show progressively decreased amount of activity as a result of adaptation. Thereafter, the direction of motion of the stimulus will change to rightward and different neurons selective for the new direction of motion will be activated. This will translate to a higher BOLD activity compared to the case when the stimulus motion direction remains the same (compare figures 6.1b and 6.1c). Such a result would indicate that direction of motion of the stimulus is being processed within this particular voxel. Similar experiments can be carried out to investigate neuronal selectivity in relation to various visual attributes such as color, orientation, shape etc. Below we describe monkey and human fMRI studies that have used the adaptation paradigm to investigate the processing visual information by neural populations.

The underlying neuronal mechanism of adaptation of the BOLD signal is currently not known. Neuronal adaptation is ubiquitously expressed in the properties of single neurons throughout the nervous system, where the firing rate of neurons in a wide range of species typically decreases during continuous presentation of a stimulus. For instance, the firing rate of neurons processing motion show adaptation during the presentation of a moving stimulus (Barlow and Hill, 1963; Ibbotson et al., 1998; Maddess et al., 1988; Oyster et al., 1972; Vautin and Berkley, 1977) (von der Heydt al., 1978; Lisberger and Movshon, 1999). Future experiments combining electrophysiological recordings and BOLD imaging will be crucial to elucidate the relationship between adaptation of the BOLD signal and neuronal activity.

6.3 Monkey fMRI Studies

fMRI in monkeys has several advantages including the ability to combine electrophysiological recordings with BOLD measurements. In addition, monkey fMRI can be used in combination with other invasive techniques like pharmacological manipulations and lesion studies. Recently, high-spatial resolution (nanoliter voxel volume) temporally resolved (as low as 20 msec segment acquisition times) fMRI has been developed for monkeys (Logothetis et al., 1999) and fMRI adaptation experiments have been designed to investigate motion processing (Tolias et al., 2001).

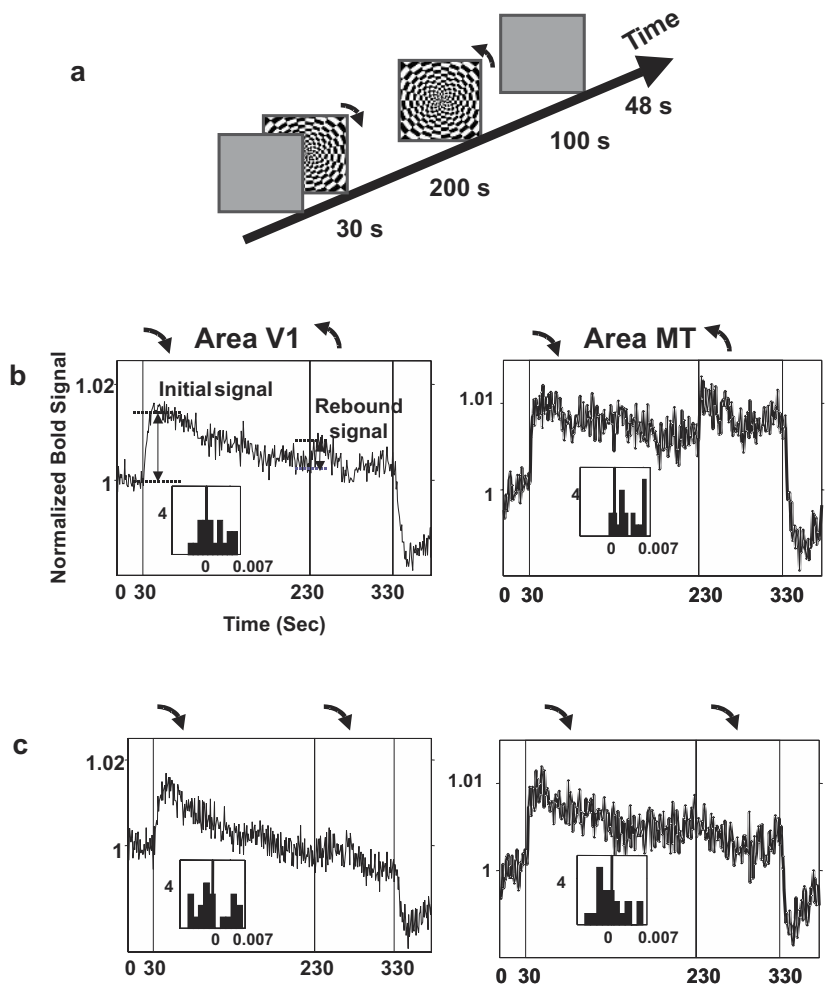


Figure 6.2 Adapted from Tolias et al., 2001: (a) Visual stimulation, see section 6.3.1. (b) Average time course of the BOLD signal in area V1 from a single slice (170 voxels, 20 repetitions). The signal is normalized to the baseline (dividing by mean activity during the initial off-condition) – a signal of 1.02 represents an increase of 2% from baseline. (c) Same analysis as (b) but with no change in the direction of motion. The mean of the distribution is not significantly different from zero (two-tailed paired t-test, $p>0.7$ for V1 and two-tailed paired t-test, $p>0.8$ for area MT). The histograms inserted show the distribution of the reactivation of the signal across all activated V1 voxels for 20 identical stimulation trials. Description of the time courses, see section 6.3.2

The methods and the measured data of the adaptation paradigm study (Tolias et al., 2001) are displayed in figure 6.2 and described in the two following sections.

6.3.1 Data Acquisition and Visual Adaptation Protocol

Two to three selected horizontal brain slices were imaged every second. Each slice had a field of view of 128x128 voxels, with each voxel size $1 \times 1mm^2$ in plane resolution and 2 mm slice thickness. The used stimulus is depicted in figure 6.2a: Thirty s of the off-condition were followed by 200 s of the ON-condition while the polar was rotating continuously in only one of the two motion directions, i.e. either clockwise or counterclockwise. In interleaved trials the direction of motion of the polar was reversed after 200 s of stimulus onset and presentation continued for another 100 s. At the point of transition in both cases the phase of rotation of the polar was reset which was visible as a fast transient without affecting the BOLD. Finally, 48 s of the off-condition was presented.

6.3.2 Signal Time Courses Observed in Visual Areas

The typical time course of BOLD activity from area V1 is illustrated for a single slice in figure 6.2 (b and c, left). The first 30 s show activity during off-condition. Clockwise and counterclockwise arrows represent ON-conditions, with the polar rotating counterclockwise and clockwise. The reactivation for each stimulation trial was defined to be the difference between the mean signal for 30 s before and after the change in the direction of motion across all activated voxels from V1. The mean of the distribution is significantly bigger than zero (two-tailed paired t-test, p¡0.05) indicating an increase in the BOLD signal after the change in the direction of motion. For areas MT in figure 6.2 (b and c, right) the mean of the distribution is significantly greater than zero (two-tailed paired t-test, p<10-4) when a change in the direction of rotation occurs.

6.3.3 Results and Interpretation

Using an fMRI adaptation paradigm a distributed network of visual areas has been identified within the visual cortex of the macaque that processes information about direction of motion of a stimulus.

The BOLD signal rises quickly to a peak after the onset of the stimulus, and then adapts slowly while the polar stimulus is rotating in the initial direction of motion. Following the reversal of the direction of motion of the stimulus, a second peak is seen in the BOLD signal (rebound signal) that reflects release from adaptation to the initial direction of motion of the polar stimulus (figure 6.2b, left). The existence of this second peak demonstrates explicitly that direction of motion information is reflected in the BOLD signal. In the control condition the direction of motion of the visual stimulus did not change (figure 6.2c, left). In both the experiment and the control conditions, the rotating polar was reset to its starting position 200 s after

the onset of the visual stimulus. This stimulus transient did not influence the time course of the BOLD signal (figure 6.2c). An even stronger release from adaptation was found in area MT (figure 6.2b; right panel) consistent with the crucial role it plays in the processing of information about the direction of motion of a stimulus (Albright, 1984; Maunsell and Van Essen, 1983; Newsome et al., 1989; Newsome and Salzman, 1993; Salzman and Newsome, 1994).

In addition, other visual areas (V2, V3, V3A and V4) were also found to participate in the processing of the direction of motion (figure 6.3a). The relative strength of the rebound- versus the initial-signal varied across these areas, indicating the difference in processing of motion signals among them (figure 6.3b).

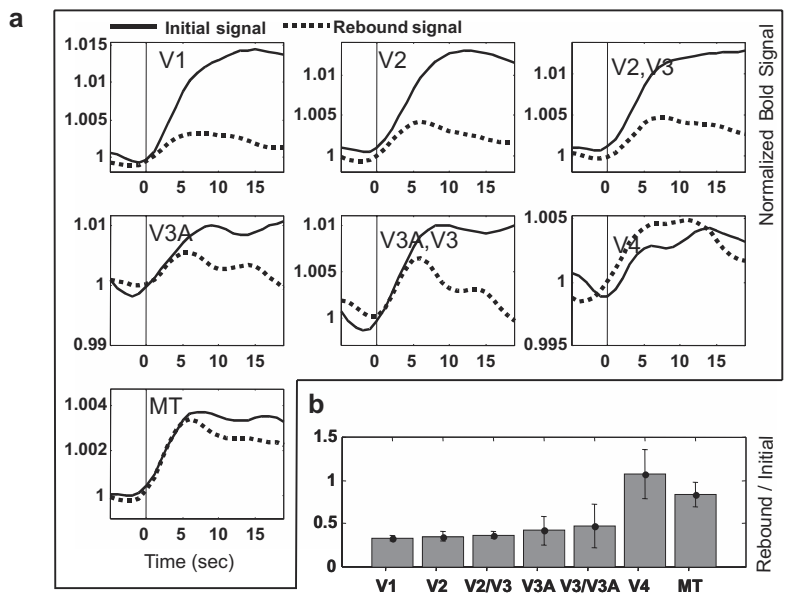


Figure 6.3 Adapted from Tolias et al., 2001. Information about direction of motion across different visual areas (a) The mean initial and rebound filtered responses (digital low pass Butterworth 8 order filter and cutoff frequency 0.125 Hz) for different visual areas are plotted in solid and dotted lines, respectively. These signals represent the mean across all slices and experimental sessions. If the mean activity (from 11 to 20 s after stimulus onset) from an individual slice for a particular visual area was less that two standard deviations above the mean of the baseline (activity during 30 s of background stimulation) then this response was excluded from the analysis. (b) BOLD directionality indices (r = Rebound-response/Initial-response). For V1 mean r = 0.33 (standard error of mean; SEM=0.03), for V2 mean r = 0.35 (SEM=0.06), for V2/V3 mean r = 0.37 (SEM=0.04), for V3A mean r = 0.42 (SEM=0.16), for V3/V3A mean r = 0.43 (SEM=0.25), for V4 mean r = 1 (SEM=0.29), for MT mean r = 0.84 (SEM=0.15). The BOLD directional index in V1 was significantly smaller from both the index of MT and V4 (two-tailed paired t-test, $p<0.001$ for V1, MT comparison and two-tailed paired t-test, $p<0.001$ for V1, V4 comparison). The BOLD directional indices for V4 and MT were not significantly different (two-tailed paired t-test, $p>0.05$).

6.3.4 Relations to Results from Single-Unit Electrophysiology

Monkey fMRI experiments allow semi-quantitative comparisons between known results from electrophysiology and the BOLD signal. When such comparisons were made, the strength of the directional selectivity reflected in the BOLD signal in some visual areas such as V1 and V4 was found to be greater from an estimate based on the number of strongly directional selective cells found in these areas. A hypothesis has been proposed that could account for the apparent difference between the single unit electrophysiology and BOLD results (Tolias et al., 2001) that is based on the numerous connections that exist among neurons within and between brain areas. According to this hypothesis, neuronal selectivity is a function of the state of adaptation. In this case neurons that under classical investigation may not be directionally selective can manifest directional selectivity after adapting to directional stimuli. To demonstrate the principle behind this hypothesis, consider a network of neurons where directionally selective cells activate other cells, so that the latter group of neurons is classically non-directionally selective. This can be achieved if heterogeneous populations of directionally selective cells converge to provide balanced input to the non-directionally selective neurons (figure 6.4a). The directionally selective cells in figure 6.4, show robust activity only when the preferred stimulus is presented in the visual field (figure 6.4a). On the other hand, the non-directional selective cells are activated when stimulated with either direction of motion. However, after the network adapts to a particular direction of motion of the visual stimulus, a change in the direction of motion will result in an increase in the activity of the "non-directionally selective" neurons, thereby contributing towards the rebound signal (figure 6.4b). Therefore, one possible source for the higher than expected BOLD directional index in visual areas such as V1 may be the different methodologies employed between the imaging and electrophysiology studies. In the fMRI study adaptation was used to probe visual selectivity whereas the electrophysiology experiments were carried out using standard selectivity mapping procedures. Although there is no way to be absolutely sure that the relationships between BOLD and known electrophysiology are not simply due to differences in the transfer function between neuronal activity and hemodynamic response in different visual areas, the adaptation-dependent selectivity hypothesis can easily be tested with standard electrophysiology. In fact, recent electrophysiology work provides evidence supporting the above hypothesis. Specifically, the orientation tuning of complex cells in macaque V1 shifts as a function of the orientation of an adapting stimulus (Muller et al., 1999; Dragoi et al., 2000).

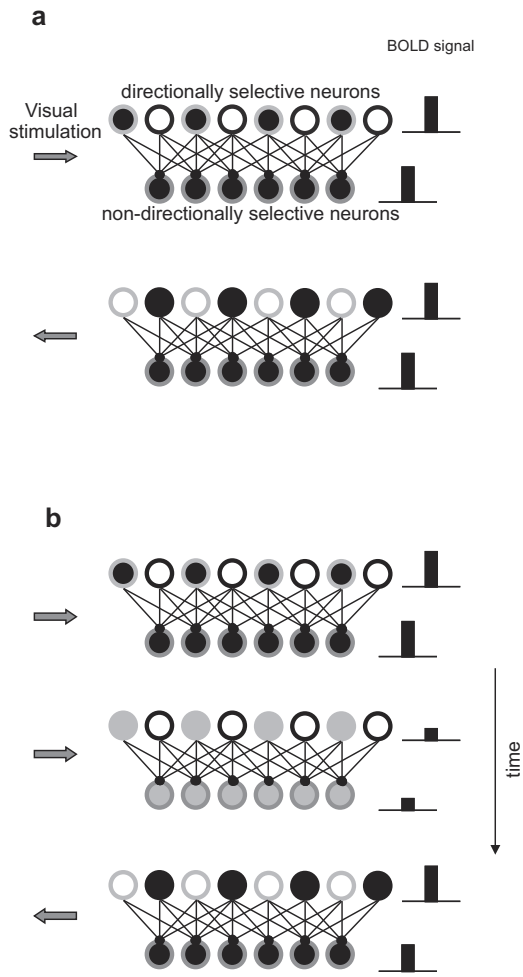


Figure 6.4 Model of proposed interactions among neurons in a network, during adaptation, processing information about direction of motion. (a) Hypothetical classical direction selectivity experiments. Upper layer black and gray outlined circles represent neurons which exhibit direction selectivity (e.g. MT cells) in the classical sense. Lower level gray outlined circles represent neurons that do not exhibit classical direction selectivity (e.g. most neurons in V1). The directionally selective neurons respond only when the preferred direction of motion (leftward or rightward pointing arrow) is presented (black filled circles). In contrast, the non-directionally selective cells respond equally during both visual stimulation conditions. (c) The BOLD signals originating from both the directionally and non-directionally selective neurons show an increase following a change in the direction of motion of the visual stimulus. This is because of the input the non-directionally selective cells receive from the directionally selective ones. One plausible implementation of such a model could come about through feedback interactions, i.e. connections between areas MT and V1. Other schemes including elaboration of local neuronal circuits are also possible.

Neuronal adaptation is a fundamental principle of single cell physiology (Barlow, 1972). The extensive connectivity between neurons in a distributed hierarchical network may mediate a change of neuronal specificity in early visual areas as a function of adaptation to high level visual attributes computed in the higher areas. Neurons in early visual areas, currently thought to lack information about certain attributes of the visual scene when tested classically, might nevertheless be able to adapt specifically to those attributes. Thereby, neurons in early visual areas may be recruited to encode changes along specific stimulus dimensions.

Recently, Kourtzi et al. (2001) investigated in monkeys by the same fMRI adaptation paradigm as above how local image features are integrated into configurations that may represent visual shapes. The adapting stimulus consisted of a rectangular area filled with randomly oriented line segments (noise stimulus) followed by one of three test stimuli: a) a pattern identical to the adapting stimulus, b) a second noise pattern where 1/3 of the line segments changed orientation randomly from the original adapting noise pattern, and c) a pattern where the same line segments changed orientation to form a collinear shape. Higher increased rebound signal during the test phase for the collinear shape than the noise pattern identified visual areas with neurons that are selective for the global configuration of shapes.

6.4 Human fMRI Studies

Human imaging studies have also employed the adaptation paradigm to define the functional neural properties of human brain areas involved in the processing of the visual input. Specifically, recent human imaging studies tested whether the neural populations in the early visual areas are tuned to visual features, such as orientation (Boynton, 2001; Ress and Heeger, 2001) and direction of motion (Huk and Heeger, 2001). To this end, observers were presented with gratings at a specific orientation or motion direction. After exposure to this adapting stimulus, observers were tested with the same stimulus in the same or in an orthogonal orientation or motion direction. Decreased fMRI responses were observed in V1 and MT/MST when the test stimuli were in the same orientation or motion direction as the adapting stimulus, respectively. However, recovery from this adaptation effect was observed for stimuli presented at an orthogonal orientation or direction of motion. These studies suggest that the neural populations in V1 and MT/MST are tuned to orientation and direction of motion, respectively. Similarly, recent studies have shown stronger adaptation in MT/MST for coherently moving plaid stimuli than for transparently moving gradings. These findings provide evidence that fMRI adaptation responses are linked to the activity of pattern-motion rather than component-motion cells in MT/MST (Huk and Heeger, 2001). Finally, selective fMRI adaptation to color contrast has been reported in V1 (Engel and Furmanski, 2001). Thus, these studies provide evidence that the fMRI signal may reveal neural selectivity in human visual regions similar to the selectivity established with neurophysiological methods.

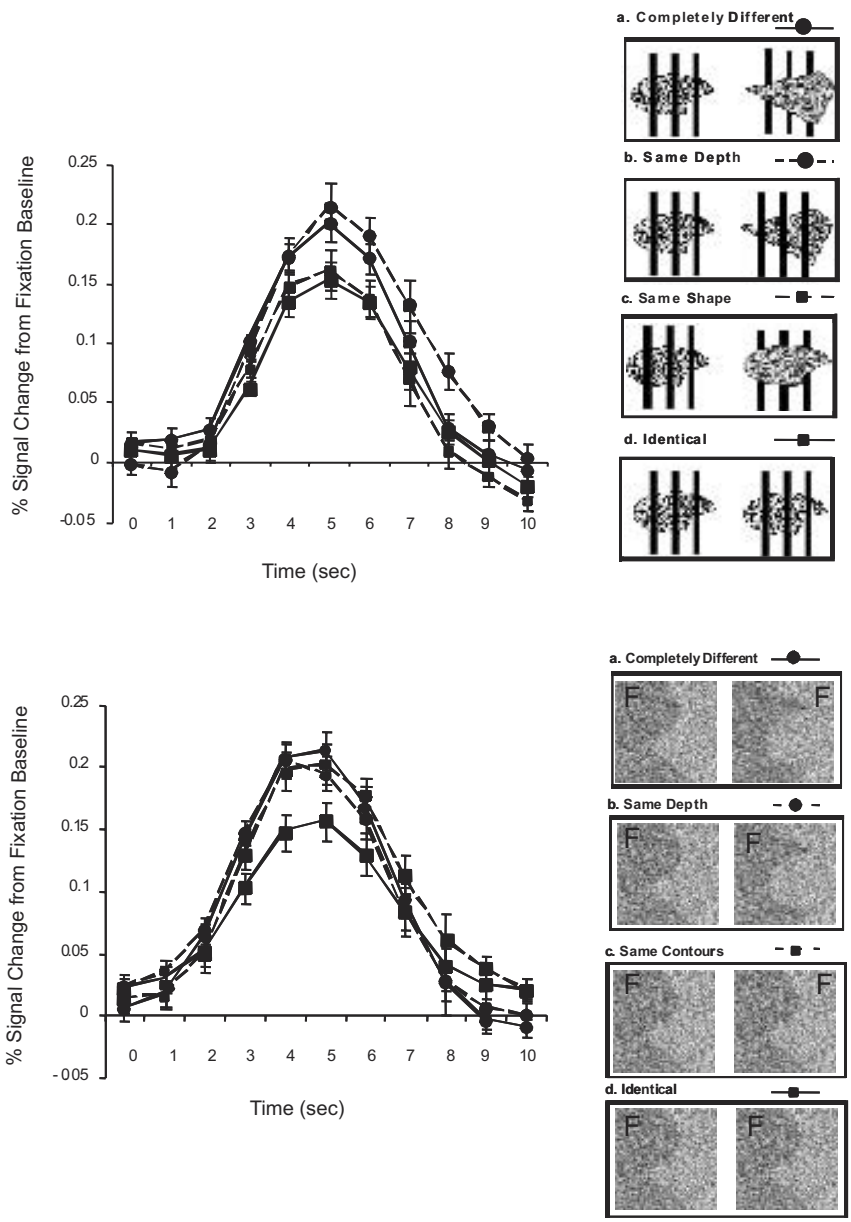
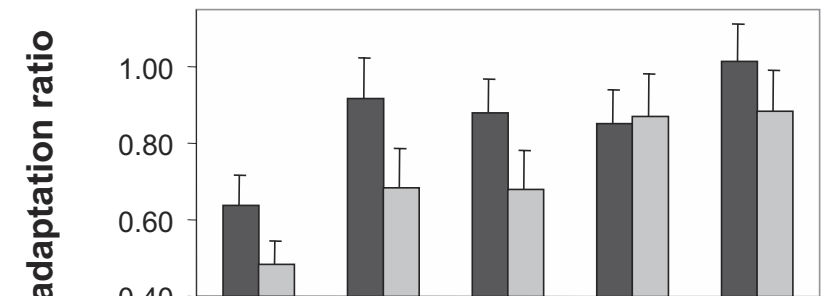


Figure 6.5 Adapted from Kourtzi and Kanwisher, 2001. Shape Processing in the LOC: Data averaged across 10 subjects showing adaptation effects in the LOC, that is decreased responses for identical images of objects (compared to the responses for different objects in a trial. Adaptation is shown (upper panel) for images that have the same perceived shape but different contours due to occlusion. In contrast, no adaptation is shown (lower panel) for images that when rendered stereoscopically have the same contours but different perceived shape due to figure ground reversal (F indicates the shape perceived as the figure in front of the background for each image).

Another set of human fMRI studies tested for selectivity to shape in higher visual areas, which are thought to be involved in the processing of visual objects. In particular, the lateral occipital complex (LOC), a region in the lateral occipital cortex extending anterior in the temporal cortex, has been shown to be involved in shape processing (Kanwisher et al., 1996; Malach et al., 1995). Recent human fMRI studies have used adaptation to test whether neural populations in the LOC show selectivity to visual properties of objects or whether they represent objects independent of image changes. Adaptation across a change between two shapes provides evidence for a common neural representation invariant to that change, while recovery from adaptation suggests neural representations selective for specific shape properties.

In particular, fMRI adaptation was used to test whether the LOC is involved in the processing of object shape independent of low level image features that define the shape (figure 6.5; Kourtzi and Kanwisher, 2001). An event-related fMRI adaptation paradigm was employed, in which a pair of consecutively-presented stimuli was presented in each trial that lasted for 3 s. These studies showed adaptation in the LOC when the perceived shape was identical but the image contours differed



	Ident	Trans	Size	Illum	View
LO	0.64	0.92	0.88	0.85	1.01
LOa/pFs	0.48	0.69	0.68	0.87	0.88

Figure 6.6 Adapted from Grill-Spector et al., 1999. Adaptation ratios are calculated for each condition by dividing the mean signal in an epoch of that condition by the mean signal in the different epoch consisting of different exemplars shown under the same viewing conditions. A ratio of 1.0 indicates no adaptation. Ratios that were significantly less than 1.0, indicate significant adaptation ($p<0.01$) and are marked by asterisks. Error bars indicate one standard error of the mean (SEM). Note that both in LO (posterior part of the LOC) and pFS/LOa (anterior part of the LOC) there is adaptation due to repetitions of identical images. However, LO and LOa/pFs exhibit different levels of adaptation especially in the translation and size epochs (for details see Grill-Spector et al., 1999). Abbreviations: Ident: repetitions of identical pictures. Trans: the same object translated in the image plane. Size: the same object shown in different sizes. Illum: same object illuminated from five different directions. View: same object depicted from different viewing angles around the vertical axis.

(because occluding bars occurred in front of the shape in one stimulus and behind the shape in the other). In contrast, release from adaptation was observed when the contours were identical but the perceived shapes were different (because of a figure-ground reversal). Consistent with these results, adaptation was also shown for grayscale images and line drawings of the same objects (Kourtzi and Kanwisher, 2000). These results suggest that neural populations in the LOC may not represent simple image features, such as contours, but higher-level shape information independent of image cues (i.e. shading and line contours).

Another human imaging study tested the effect of different stimulus transformations, namely position, size, orientation, and illumination change, on the BOLD signal in the LOC (figure 6.6; (Grill-Spector et al., 1999; Grill-Spector and Malach, 2001). Adaptation was observed when the observers were presented repeatedly with identical images of objects. Stronger recovery from adaptation was shown across orientation or illumination changes compared to size and position changes. Interestingly, adaptation effects across orientation and size changes were observed more strongly in the anterior rather than the posterior regions of the LOC.

A particularly interesting aspect of repetition and adaptation is their relationships to learning. Recent studies have used adaptation to test whether learning and stimulus familiarity are associated with changes of the BOLD signal. Specifically, event-related fMRI studies have shown decreased activation in temporal and frontal areas for repeated presentation of objects (Buckner et al., 1998). This repetition suppression effect has been observed for familiar (Henson et al., 2000) namable objects (James et al., 1999) and it has been reported to be rather long lasting (e.g. 3 days) (van Turennout et al., 2000). It has been suggested that this adaptation effect is related to a psychophysical effect known as visual priming in which repeated presentation of a stimulus results in faster and more accurate observer performance in visual discrimination or object naming tasks (Schacter and Buckner, 1998; Wiggs and Martin, 1998). Neurophysiological studies (Miller et al., 1991; Li et al., 1993) have also observed this repetition suppression effect and have proposed that it may reflect signals from neural populations that become smaller but more highly tuned to specific shape properties after the repeated presentations of objects (Desimone, 1996). As a result these neural populations become more selective to the repeated stimuli and may support more efficient behavioral responses. Finally, adaptation effects have been observed in higher cognitive tasks other than visual processing, such as semantic classification of objects (Buckner and Koutstaal, 1998) and procedural motor learning (Karni et al., 1995). Taking these results together we can conclude that the fMRI adaptation technique has been proven to be a useful tool for defining the functional properties of human brain regions involved in visual analysis and for investigating the representations of visual features and shapes that may mediate higher cognitive processes.

6.5 Discussion

Adaptation is a powerful tool for studying the properties of networks of neurons with imaging techniques. Specifically, in this chapter we present evidence that adaptation paradigms can be used in imaging experiments to characterize properties of neuronal populations beyond the spatial resolution of current imaging techniques. The validity of the adaptation technique is illustrated by results from monkey and human fMRI studies. In particular, the adaptation experiments in monkeys showed the existence of strong selectivity for information about the direction of motion in area MT. These results are consistent with previous work using single unit recording, microstimulation and lesion techniques showing the crucial role MT plays in the processing and perception of stimulus motion direction (Albright, 1984; Maunsell and Van Essen, 1983; Newsome et al., 1989; Newsome and Salzman, 1993; Salzman and Newsome, 1994).

In addition, adaptation can reveal certain response properties of neurons beyond those known from standard neuronal selectivity experiments. Unfortunately, the relationship between the adaptation of the BOLD signal and neuronal activity is currently not known. Recently, simultaneous recording of BOLD and electrophysiological signals using microelectrodes have become possible (Logothetis et al., 2001). Recording simultaneously the BOLD signal and electrophysiological activity during adaptation is likely to provide further insights about the relationship between BOLD and neuronal adaptation.

Acknowledgments

We thank Natasha Sigala and Stelios M. Smirnakis for critically reading the manuscript.

References

T. D. Albright. Direction and orientation selectivity of neurons in visual area MT of the macaque. Journal of Neurophysiology 52, 1106-1130, 1984.

H. B. Barlow. Single units and sensation: a neuron doctrine for perceptual psychology? Perception 1, 371-394, 1972.

H. B. Barlow and R. M. Hill. Evidence for a physiological explanation of the waterfall phenomenon and figural after effects. Nature 200, 1345-1347, 1963.

G. M. Boynton. Orientation-specific pattern adaptation measured with event-related fMRI. Vision Sciences Society Abstracts, 25, 2001.

J. W. Belliveau, M. S. Cohen, R. M. Weisskoff, B. R. Buchbinder, and B. R. Rosen. Functional studies of the human brain using high-speed magnetic resonance imaging. J. Neuroimaging 1, 36-41, 1991.

R. L. Buckner, J. Goodman, M. Burock, M. Rotte, W. Koutstaal, D. Schacter, B. R. Rosen, and A. M. Dale. Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. Neuron 20, 285-296, 1998.

R. L. Buckner and W. Koutstaal. Functional neuroimaging studies of encoding, priming, and explicit memory retrieval. Proc. Natl. Acad. Sci. U. S. A 95, 891-898, 1998.

P. S. Churchland and T. J. Sejnowski. Perspectives on Cognitive Neuroscience. Science 741-745, 1988.

R. Desimone. Neural mechanisms for visual memory and their role in attention. Proc. Natl. Acad. Sci. U. S. A 93, 13494-13499, 1996.

V. Dragoi, J. Sharma, and M. Sur. Adaptation-induced plasticity of orientation tuning in adult visual cortex. Neuron 28, 287-298, 2000.

S. A. Engel and C. S. Furmanski. Selective adaptation to color contrast in human primary visual cortex. J. Neurosci 21, 3949-3954, 2001.

K. Grill-Spector, T. Kushnir, S. Edelman, G. Avidan, Y. Itzchak, and R. Malach. Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 24 , 187-203, 1999.

K. Grill-Spector and R. Malach. fMRI-Adaptation: a Tool for Studying the Functional Properties of Human Cortical Neurons. Acta Psychologica (in press).

R. Henson, T. Shallice, and R. Dolan. Neuroimaging evidence for dissociable forms of repetition priming. Science 287, 1269-1272, 2000.

A. C. Huk, D. Ress, and D. J. Heeger. Neuronal basis of the motion aftereffect reconsidered. Neuron 2001, 32, 161-72, 2001.

A. C. Huk and D. J. Heeger. Pattern-motion responses in human visual cortex. Nat Neurosci. (in press).

M. R. Ibbotson, C. W. Clifford, and R. F. Mark. Adaptation to visual motion in directional neurons of the nucleus of the optic tract. J. Neurophysiol. 79, 1481-1493, 1998.

T. W. James, G. K. Humphrey, J. S. Gati, R. S. Menon, and M. A. Goodale. Repetition priming and the time course of object recognition:an fMRI study. NeuroReport 10, 1019-1023, 1999.

N. Kanwisher, M. M. Chun, J. McDermott, and P. J. Ledden. Functional imagining of human visual recognition. Brain Res. Cogn Brain Res. 5, 55-67, 1996.

A. Karni, G. Meyer, P. Jezzard, M. M. Adams, R. Turner, and L. G. Ungerleider. Functional mri evidence for adult motor cortex plasticity during motor skill learning. Nature 377, 155-158, 1995.

D. S. Kim, T. Q. Duong, and S. G. Kim. High-resolution mapping of iso-orientation columns by fMRI. Nat. Neurosci 3, 164-169, 2000a.

D. S. Kim, T. Q. Duong, and S. G. Kim. Reply to "Can current fMRI techniques reveal the micro-architecture of cortex?". Nat. Neurosci 3, 414, 2000b.

Z. Kourtzi and N. Kanwisher. Cortical regions involved in perceiving object shape. J. Neurosci. 20, 3310-3318, 2000.

Z. Kourtzi and N. Kanwisher. The human lateral occipital complex represents perceived object shape. Science, 24, 1506-9, 2001.

Z. Kourtzi, A. S. Tolias, B. Prause, M. Augath, T. Trinath, and N. K. Logothetis. fMRI adaptation for visual forms in the monkey brain. Society for Neuroscience, San Diego, 2001.

L. Li, E. K. Miller, and R. Desimone. The representation of stimulus familiarity in anterior inferior temporal cortex. Journal of Neurophysiology 69, 1918-29, 1993.

S. G. Lisberger, and J. A. Movshon. Visual motion analysis for pursuit eye movements in area MT of macaque monkeys. Journal of Neuroscience 19, 2224-2246, 1999.

N. K. Logothetis. Can current fMRI techniques reveal the micro-architecture of cortex? Nat. Neurosci 3, 413-414, 2000.

N. K. Logothetis, H. Guggenberger, S. Peled, and J. Pauls. Functional imaging of the monkey brain [see comments]. Nat. Neurosci. 2, 555-562, 1999.

N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological Investigation of the Basis of the fMRI Signal. Nature (in press).

T. Maddess, M. E. McCourt, B. Blakeslee, and R. B. Cunningham. Factors governing the adaptation of cells in area-17 of the cat visual cortex. Biol. Cybern. 59, 229-236, 1988.

R. Malach, J. B. Reppas, R. R. Benson, K. K. Kwong, H. Jiang, W. A. Kennedy, P. J. Ledden, T. J. Brady, B. R. Rosen, and R. B. Tootell. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proceedings of the National Academy of Sciences of the United States of America 92, 8135-8139, 1995.

J. H. R. Maunsell and D. C. Van Essen. The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. J. Neurosci. 3, 2563-2586, 1983.

E. K. Miller, P. M. Gochin, and C. G. Gross. Habituation-like decrease in the responses of neurons in inferior temporal cortex of the macaque. Visual Neuroscience 7, 357-362, 1991.

J. R. Muller, A. B. Metha, J. Krauskopf, and P. Lennie. Rapid adaptation in visual cortex to the structure of images. Science 285, 1405-1408.

W. T. Newsome, K. H. Britten, and J. A. Movshon. Neuronal correlates of a perceptual decision. Nature 341, 52-54, 1999.

W. T. Newsome and C. D. Salzman. The neuronal basis of motion perception. Ciba Foundation Symposium 174, 217-30. Oyster,C.W., Takahashi,E., and Collewijn,H. (1972). Direction-selective retinal ganglion cells and control of optokinetic nystagmus in the rabbit. Vision Res. 12, 183-193, 1993.

D. Ress and D. J. Heeger. fMRI measurements of orientation-selective adaptation. Vision Sciences Society Abstracts, 49, 2001.

C. D. Salzman and W. T. Newsome. Neural Mechanisms for Forming a Perceptual Decision. Science 264231, 231-237, 1994.

D. L. Schacter and R. L. Buckner. Priming and the brain. Neuron 20, 185-195, 1998.

A. S. Tolias, S. M. Smirnakis, M. Augath, T. Trinath, and N. K. Logothetis. Motion Processing in the Macaque; Revisited with Functional Magnetic Resonance Imaging. J Neurosci. 21(21). 8594-8601, 2001.

M. van Turennout, T. Ellmore, and A. Martin. Long-lasting cortical plasticity in the object naming system. Nat. Neurosci 3, 1329-1334, 2000.

R. G. Vautin and M. A. Berkley. Responses of single cells in cat visual cortex to prolonged stimulus movement: neural correlates of visual aftereffects. J. Neurophysiol. 40, 1051-1065, 1977.

R. von der Heydt, P. Hanny, and C. Adorjani. Movement aftereffects in the visual cortex. Arch. Ital. Biol. 116, 248-254, 1978.

C. L. Wiggs and A. Martin. Properties and mechanisms of perceptual priming. Curr. Opin. Neurobiol. 8, 227-233, 1998.

This page intentionally left blank

II EGM/MEG Data Analysis

This page intentionally left blank

7 Independent Components of Magnetoencephalography: Localization

Akaysha C. Tang
Departments of Psychology, Computer Science, and Neurosciences
University of New Mexico, Albuquerque, NM 87131, USA
akaysha@unm.edu

Barak A. Pearlmutter
Departments of Computer Science, Neurosciences, and EECE
University of New Mexico, Albuquerque, NM 87131, USA
bap@cs.unm.edu

We applied second-order blind identification (SOBI), an independent component analysis (ICA) method, to MEG data collected during cognitive tasks. We explored SOBI's ability to help isolate underlying neuronal sources with relatively poor signal-to-noise ratios, allowing their identification and localization. We compare localization of the SOBI-separated components to localization from unprocessed sensor signals, using an equivalent current dipole (ECD) modeling method. For visual and somatosensory modalities, SOBI preprocessing resulted in components that can be localized to physiologically and anatomically meaningful locations. Furthermore, this preprocessing allowed the detection of neuronal source activations that were otherwise undetectable. This increased probability of neuronal source detection and localization can be particularly beneficial for MEG studies of higher level cognitive functions, which often have greater signal variability and degraded signal-to-noise ratios than sensory activation tasks.

7.1 Introduction

Magnetoencephalography (MEG) is a passive functional brain imaging technique which, under ideal conditions, can monitor the activation of a neuronal population with a spatial resolution of a few mm and with millisecond temporal resolution (Hämäläinen et al., 1993; George et al., 1995). Typical signals associated with

neuronal activity are on the order of one hundred fT, while the noise signals within a shielded room tend to be much larger (Lewine and Orrison, 1995). Furthermore, the intrinsic sensor noise is comparable in magnitude to small neuronal signals. Therefore, what the sensors record during an experiment is always a mixture of small neuromagnetic and large noise signals. This relatively poor signal-to-noise ratio[1] can affect the localization of neuronal activity.

Several independent component analysis (ICA) algorithms, such as second-order blind identification (SOBI) (Belouchrani et al., 1993; Cardoso, 1994), Infomax (Bell and Sejnowski, 1995), and fICA (Hyvärinen and Oja, 1997), have been applied to EEG data (Makeig et al., 1996, 1997, 1999b; Jung et al., 2000a,b) and MEG data (Vigário et al., 1998; Tang et al., 2000a; Vigário et al., 1999, 2000; Wübbeler et al., 2000; Ziehe et al., 2000; Cao et al., 2000). In both applications, ICA methods have proven useful for artifact removal and for improving the signal-to-noise ratio (Jung et al., 2000a,b; Vigário et al., 1998; Tang et al., 2000a). For general reviews of ICA see Amari and Cichocki (1998); Cardoso (1998); Hyvärinen (1999); Vigário et al. (2000).

For MEG, in addition to separating various noise signals from the neuromagnetic signals, SOBI and fICA have been shown to separate one neuronal source from another between and within the same modality (Tang et al., 2000a; Vigário et al., 1999, 2000). To localize functionally independent neuronal sources or to simultaneously localize and recover the time course of these neuronal sources, a variety of algorithms have been proposed (Mosher et al., 1992; Kinouchi et al., 1996; Sekihara et al., 1997; Nagano et al., 1998; Mosher and Leahy, 1998; Uutela et al., 1998; Mosher and Leahy, 1999; Schwartz et al., 1999; Sekihara et al., 2000; Huang et al., 2000; Aine et al., 2000; Ermer et al., 2000; Cao et al., 2000; Schmidt et al., 1999). Given their capability to separate noise and neuronal signals, ICA algorithms are expected to benefit all source localization methods by providing them with input signals that are more likely to be associated with functionally independent neuronal sources.

It was found, however, that the fICA-separated components yielded localization results qualitatively similar to those arrived at without ICA preprocessing (Vigário et al., 1999). Consequently, no substantial benefits from ICA were reported for neuromagnetic source localization. As one of the strengths of ICA is its ability to separate noise from the signals of interests, whether ICA could offer any advantage in source localization should depend on the signal-to-noise ratio in the sensor data. The experiment reported by Vigário et al. (1999) was optimally designed to produce strong and focal activation of a small number of neuromagnetic sources, and therefore has high signal-to-noise ratios. Under such optimal conditions, ICA could

1. Unless otherwise indicated, we use signal-to-noise ratio in the sense defined in signal detection theory. Signals refer to the neuromagnetic signal of interest. Noise refers to all other signals including environmental and sensor noise and other background brain signals.

not improve much upon the already good localization provided by conventional methods.

In this chapter, we applied ICA to neuromagnetic signals with relatively poor signal-to-noise ratios collected during cognitive tasks involving large trial-to-trial variability in neuronal source activation and from a much larger number of sources. We localized these neuronal sources using the equivalent current dipole (ECD) modeling method (Neuromag) on SOBI-separated components, and on unprocessed sensor data. We found that SOBI preprocessing resulted in the localization of neuronal sources that could not be found when the dipole fitting method was directly applied to the sensor data. In addition, the process of localizing separated components required significantly less subjective judgment regarding which sensors to exclude from the analysis[2] and at what time the dipoles are fitted. We suggest that ICA methods can be particularly effective and efficient in the study of higher level cognitive functions when the neuronal source activations are often characterized by their greater degree of variability and lower signal-to-noise ratios.

7.2 Methods

7.2.1 Cognitive Tasks

We collected MEG data from four right-handed subjects (two females and two males) during four visual reaction time tasks originally designed to study temporal lobe memory functions (Tang et al., 2000b). These tasks are described in detail in section 7.5. Here, we offer a brief description. In each task, a pair of colored patterns, one of which was the target, was presented on the left and right halves of the display screen. The subject was instructed to press either the left or right button when the target appeared on the left or right, respectively. In all tasks, the target was not described to the subject prior to the experiment. The subject was to discover the target by trial and error using auditory feedback (low and high tones corresponded to correct and incorrect responses, respectively). All subjects were able to discover the rule within a few trials.

The tasks differed in the memory load required for determining which of the pair is the target. Task one served to familiarize the subjects with all visual patterns. The subjects simply viewed the stimuli and were asked to press either the left or right button at their own choice while making sure approximately equal numbers of left and right button presses were performed. As such, task one placed little memory demand on the subject. Task two involved remembering a single target pattern which appeared on each trial paired with another pattern. Subjects pressed the right or left button to indicate whether the target pattern was on the left or

2. It is a common practice to select 20–30 sensors over the brain region of interest for dipole fitting.

the right. Task three involved remembering multiple targets, each always paired with the same non-target. Task four was the most complex. In task four, targets were context sensitive in a circular fashion, as in the game rock-paper-scissors. The amount of cognitive processing beyond the initial sensory processing increased successively from task one to task four.

We used data from these complex cognitive tasks to evaluate the capability of SOBI (see section 7.6.1) because of the relatively poor signal-to-noise ratios involved in comparison to sensory activation tasks. Specifically, these tasks involved (1) large visual field stimulation without the use of fixation points, (2) incidental somatosensory stimulation as a result of button presses during reaction time tasks, and (3) highly variable button press responses (because precisely what form of the thumb movement should be make, how the mouse was to be held, and where the hands should rest were not specified.) These sources of variability in visual and somatosensory activation can lead to poor signal-to-noise ratios in the average responses, making it particularly difficult to localize the neuronal sources from unprocessed averaged sensor data. The involvement of higher level cognitive functions, memory demands, and the small number of trials (90 in most cases) collected under each task condition, further decreased the signal-to-noise ratios in the averaged sensor data. These tasks therefore offered a set of challenging datasets in which the advantages of ICA methods could be revealed.

7.2.2 Selection of ICA Methods

In selection of ICA algorithms, one important consideration is the robustness of the algorithm to sensor noise. Instantaneous and summary algorithms are two extremes of ICA algorithms that differ in whether each point in time is considered in isolation. Instantaneous algorithms, such as Bell-Sejnowski Infomax (Bell and Sejnowski, 1995) and fICA (Hyvärinen and Oja, 1997), make repeated passes through the dataset and update the unmixing matrix in response to the data at each time point. They are derived under the assumption that the signals are white, and their results should therefore be invariant to shuffling of the data. As a consequence of this, they cannot take advantage of the temporal structure of each source as a cue for correct separation. In contrast, summary algorithms first make a pass through the data while summary statistics are accumulated by averaging; they then operate solely upon the summary statistics to find the separation matrix. Some summary algorithms collect statistic that allow them to make use of the temporal structure of the sources as a cue for separation. More importantly, summary algorithms should in general be relatively insensitive to sensor noise, because their summary statistics are averages over time. The relatively poor signal-to-noise ratios in MEG data suggested the choice of a summary algorithm rather than an instantaneous algorithm.

When it can be assumed that each source has a broad autocorrelation function, as is the case with brain signals, the summary algorithm SOBI (Belouchrani et al., 1993; Cardoso, 1994) can use this temporal structure as a cue and give high quality separation while imposing rather modest computational requirements.

SOBI extracts a large set of statistics from the dataset, which it uses for the separation. Each of these statistics is calculated by averaging across the dataset, which makes the algorithm robust against noise. The particular statistics calculated are the correlations between pairs of sensors at a fixed delay, $\langle x_i(t)\, x_j(t+\tau)\rangle$. This makes good use of abundant but noisy data, and most importantly, SOBI can be tuned by modifying its set of delays (see section 7.6), allowing its users to gently integrate a very weak form of prior knowledge, namely knowledge of the length constant of the autocorrelation function. Although Bell-Sejnowski Infomax and fICA have been previously applied to MEG and EEG data, and other ICA algorithms, such as Contextual ICA (Pearlmutter and Parra, 1996) and Sparse Decomposition (Zibulevsky and Pearlmutter, 2001) are locally available, we selected SOBI as our ICA method based on the above properties of SOBI. However, we have not conducted a systematic comparison of ICA methods for our MEG data.

7.2.3 Second-Order Blind Identification

SOBI is considered blind as it makes no assumptions about the form of the mixing process. In other words, *SOBI does not attempt to solve the inverse problem* or use the physics of the situation in any way. It does not try to estimate currents, or know about Maxwell's equation or any of its consequences. The only physical assumption made about the mixing process is that it is instantaneous and linear.

Let $\mathbf{x}(t)$ be an n-dimensional vector of sensor signals, which we assume to be an instantaneous linear mixture of n unknown independent underlying sources $s_i(t)$, *via* the unknown stationary $n \times n$ mixing matrix $\mathbf{A}$,

$$\mathbf{x}(t) = \mathbf{A}\, \mathbf{s}(t) \tag{7.1}$$

The ICA problem is to recover $\mathbf{s}(t)$, given the measurements $\mathbf{x}(t)$ and nothing else. This is accomplished by finding a matrix $\mathbf{W}$ which approximates $\mathbf{A}^{-1}$, up to permutation and scaling of its rows. SOBI assumes that the sources are statistically independent in time, and not necessarily orthogonal in space. It finds $\mathbf{W}$ by minimizing the correlation[3] between one recovered source at time t and another at time $t+\tau$.

The particular set of delays τ we used were chosen to cover a reasonably wide interval without extending beyond the support of the autocorrelation function. Measured in units of samples, at our 300 Hz sampling rate, the delays[4] were

3. For justification for this minimization, see Discussion.
4. The choice of delays can affect the results of separation. Depending on the types of sources activated by the behavioral task, the selection of delays can have complex interactions with the latency of evoked responses. This is an important topic and deserves a separate study.

$$\tau \in \{\ 1,2,3,4,5,6,7,8,9,$$
$$10,12,14,16,18,$$
$$20,25,30,35,40,45,50,55,$$
$$60,65,70,75,80,85,90,95,100\ \}.$$

Each recovered $s_i(t)$ also has a sensor space projection that gives the sensor readings of $s_i(t)$ (see section 7.6.2). This sensor projection can be displayed as a field map, and can be used as input to source localization algorithms. For example, after calculating its sensor projection, we can repackage a component for localization by Neuromag dipole modeling tools.

SOBI shares a number of weaknesses with all ICA methods: they all assume that there are as many sensors as sources; they all make some sort of independence assumption; they all assume that the mixing process is linear; and they all assume that the mixing process is stable. See section 7.4.3 for further discussion.

7.2.4 Localization of Separated Components

SOBI was performed on continuous[5] 122-channel data collected during the entire period of the experiment, sampled at 300 Hz, and band-pass filtered at 0.03–100 Hz. It generated 122 components,[6] each a one-dimensional time series with an associated field map (see section 7.6.2). Each component potentially corresponds to a set of magnetic field generators.

Event triggered averages were calculated from their continuous single-trial time series for all 122 separated components, where the triggering events were either sensory stimuli or behavioral responses. For the specific tasks used here, there were typically 10–20 components in each experiment which showed responses locked to either stimuli or to button presses. Those with stimulus- or motor-locked responses were candidate neuronal generators, since they showed task related activation. Those with responses locked onto other external events, such as eye blinks or heart beats (detected using EOG and EKG), were considered known noise sources. The rest were treated as non-task-related noise sources.

For a task related component, if its field map and time course were consistent with known neurophysiological and neuroanatomical facts, we considered it a neuronal component reflecting the activity of a neuronal generator. For example, if the field map of a component shows activation over the occipital cortex and the visual stimulus triggered average for this component contains an evoked response that peaks between 50–100ms, then it is considered to reflect the activity of a visual source in the occipital lobe. Using this procedure, neuronal and non-neuronal

5. ICA algorithms can be applied to cross-trial averages rather than continuous data, as in Makeig et al. (1999c).
6. ICA algorithms produce the same number of components as there are channels in their input.

generators were separated and identified (Tang et al., 2000a,b). A dipole fitting method was then applied to the identified neuronal components. The input to the dipole fitting algorithm (Neuromag, xfit, least square) was the field map, and the output was the location of ECDs projected onto the subject's structural MRI images.

This same dipole fitting algorithm was used for localization with and without SOBI preprocessing. Because our goal was to evaluate whether ICA methods can improve source localization, we were not concerned with whether the least square dipole fitting was as good as more recent more sophisticated source modeling methods. Our interest was not in localization accuracy *per se*, but in the *comparative* performance of a given localization method when used alone, as opposed to being coupled with ICA.

In statistical comparison, to match the common practice in source modeling without SOBI, a subset of channels (20–30) over the region of interest were selected for dipole fitting with both methods. To localize each separated component, we chose channels over the region of interest showing stronger responses to the source. For localization without SOBI (the conventional practice) we began with the channels selected for SOBI localization, and then modified the selections to obtain a more dipolar field pattern.[7] If these modifications improved the results then we used them, otherwise we used the original channel selections. This procedure gave the conventional practice an advantage because the event-triggered-average responses were cleaner in the separated components than in the raw data. In fact, the raw data were often so noisy that no channels could have been selected by following the same procedure on the raw data, and therefore no localization could have been performed without the channel selection information enabled by SOBI.

To localize a component, we used its field map as input to the dipole fitting program.[8] One can select any time during the average time window to fit the dipole because the dipole solution for a component is invariant to time (see section 7.6.2). This independence of localization results from the dipole fitting time can significantly simplify the dipole localization process, making it less subjective than dipole localization directly from the sensor data, without the use of ICA. Using the conventional method, the time at which a dipole was fitted affects the final estimated dipole location.

To localize neuronal sources without SOBI preprocessing, we used event triggered *sensor* data (averages) as inputs to the dipole fitting program. We first chose the time with the largest evoked response amplitude within the time window of interest.

---

7. A field map is judged dipolar by visual inspection (Hari and Salmelin, 1997). If it contains two sets of concentric contour lines, the field is considered dipolar. If it contains more or less than two sets, the field is not dipolar.

8. Theoretically, one sampling point in time across all sensors contains all information about a source. In practice the Neuromag software (xfit) needs a time series of at least several samples. Therefore, we calculated the event-triggered average for the component of interest and made an input .fif file containing the average.

Then a subset of channels (20–30) over the region of interest were selected. When the contour maps were single-dipolar for the selected channels at the time chosen, a single dipole fit was performed. Otherwise multiple dipoles were fitted. For details of the process, see the xfit manual. In the examples shown in the figures, all channels were used in the dipole fitting to show that SOBI can identify dipolar sources without any channel selection.

7.3 Results

7.3.1 SOBI Decomposition: Time Courses and Sensor Projections

Using SOBI, continuous MEG signals from 122 channels were separated into 122 components. Each of these components has a time course and an associated sensor projection. The time course can be averaged across multiple trials using either the visual stimulus onset or the button press as a trigger. It can also be displayed as an *MEG image* (*e.g.* figure 7.6, right), a pseudo-colored bitmap in which the responses of a given component during an entire experiment can be parsimoniously displayed (Jung et al., 1999). Typically, each row represents one discrete trial of stimulation and multiple trials are ordered vertically from top to bottom. See section 7.6.3 for details on the process of giving sensible units to the components. As shown in the overlay plots of the visual stimulus and button press triggered averages for all 122 components (figure 7.1c,d), only a small fraction of the components showed task related responses. For clarity, these task related components are shown separately in figure 7.1a,b.

The components can be displayed in the sensor domain in field maps (figure 7.6, left) or in a fullview graph using the Neuromag software xfit (figure 7.2–7.5). The sensor projections for two components are shown: one for a visual component (figure 7.2) and the other for a sensory-motor component (figure 7.3). It is clear that the two components are projected selectively to sensors over the visual and sensory-motor cortices. For comparison, the fullview plots of the sensor projection from the raw data (mixture of all components) are shown in figure 7.4–7.5.

7.3.2 Energy in Separated Components

We divided components into the following six categories: visual, somatosensory, ocular artifacts, 60 Hz, sensor jumps, and other.[9] Visual and somatosensory components were identified by their clearly visible evoked responses in the MEG images and by their activation patterns in the field maps (see following sections). These

9. Sensor jumps refer to a peculiar property of the SQUID sensors which cause an enormous and nearly instantaneous DC shift. "Other" includes any components that do not belong to the first five categories.

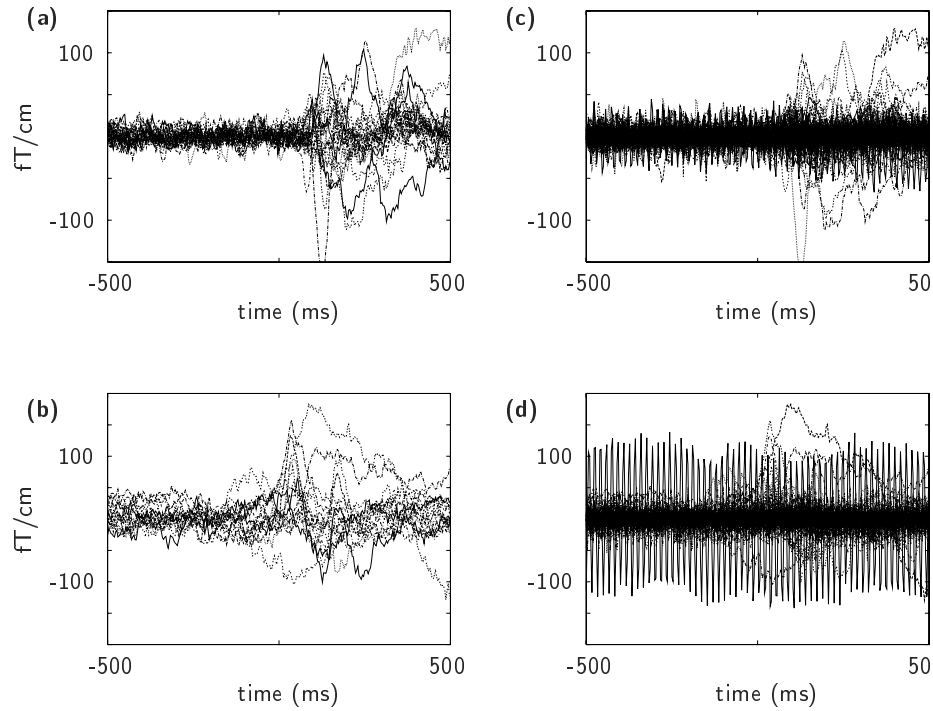


Figure 7.1 Event-triggered averages for groups of separated components ($N = 90$ trials). **(a)** Components showing visual-stimulus-triggered responses, triggered on visual stimulus onset. **(b)** Components showing button-press-triggered responses, triggered on button presses. **(c)** All components, triggered on visual stimulus onset. **(d)** All components, triggered on button presses.

neuronal components were further verified by the consistency between the response latency, shown in the MEG images, and the spatial location of sensor activation, shown in the field maps (see subsequent sections). Ocular artifact sources were identified by their characteristic activation patterns in the field map and their large amplitude responses in the MEG image (figure 7.6a), which match signals measured by EOG (not shown). The 60 Hz components were identified by the clearly visible 60 Hz cyclic activity in the MEG images in figure 7.6b (see also Tang et al., 2000a). Sensor jump components were easily identified by the single-sensor activation in the field maps and sometimes by high contrast lines or dots in the MEG images (figure 7.6c).

For these five types of identified components, we calculated the amount of energy in each (see section 7.6.4), across all subjects and all tasks, using a window of 200ms after either the visual stimulus presentation or the button presses[10] (see

10. Window specification will affect the calculated energy.

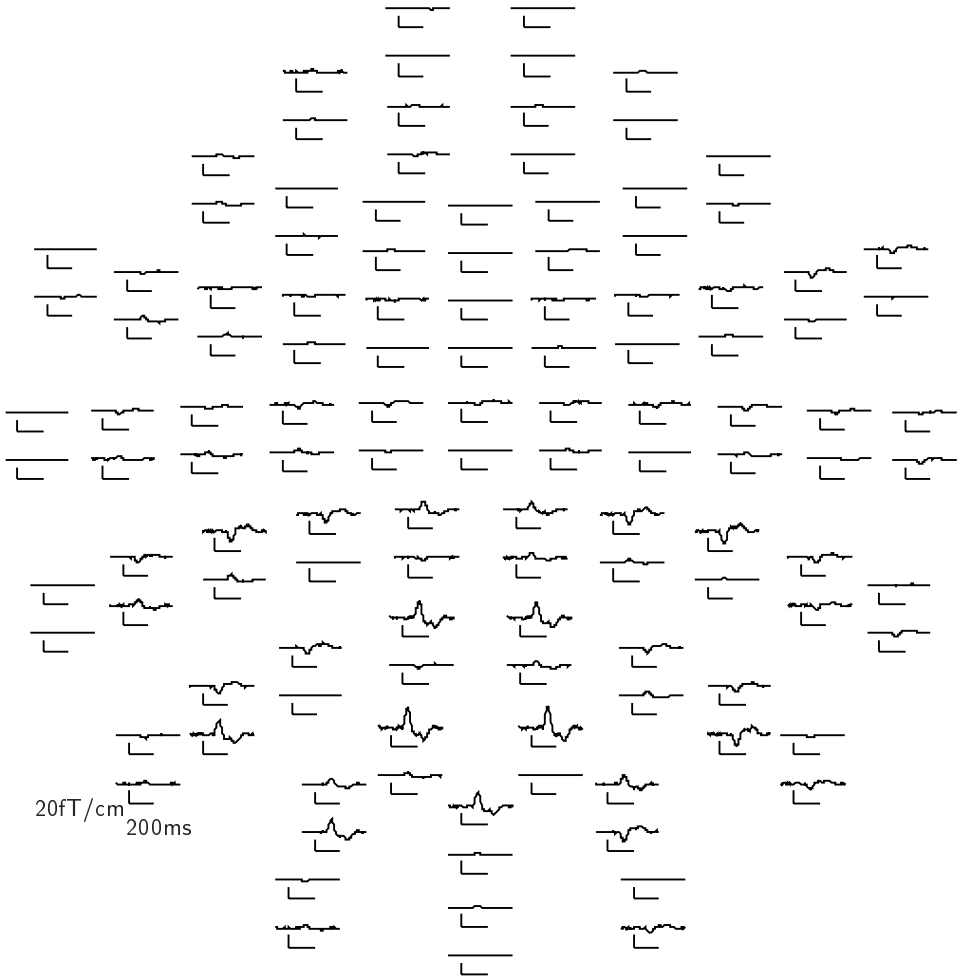


Figure 7.2 Sensor projection of component showing selective sensor activation over the occipito-parietal cortex ($N = 90$ trials, visual stimulus triggered averages). Compare with unseparated data in figure 7.4.

section 7.6.4). This window was chosen to cover all neuronal responses. The amount of energy in a single component varied widely, between 0.17% and 71% of the total energy across all sensors. This range differed among the five categories of the components (Table 7.1). The energies in the visual and somatosensory components (using visual stimuli and button presses as triggers respectively) were 10.0±1.02% ($N = 29$) and 4.65±0.74% ($N = 10$). Using button presses as triggers (because subjects tended to blink after the button press responses), the energy in the ocular artifact components were 24.86±4.67% ($N = 16$). Since both 60 Hz signals and sensor jumps were not task related, the energy in these two types of components were calculated using both visual stimulation and button presses as triggers and then averaged. The total energy in the 60 Hz sources was 10.19±1.73% ($N = 32$)

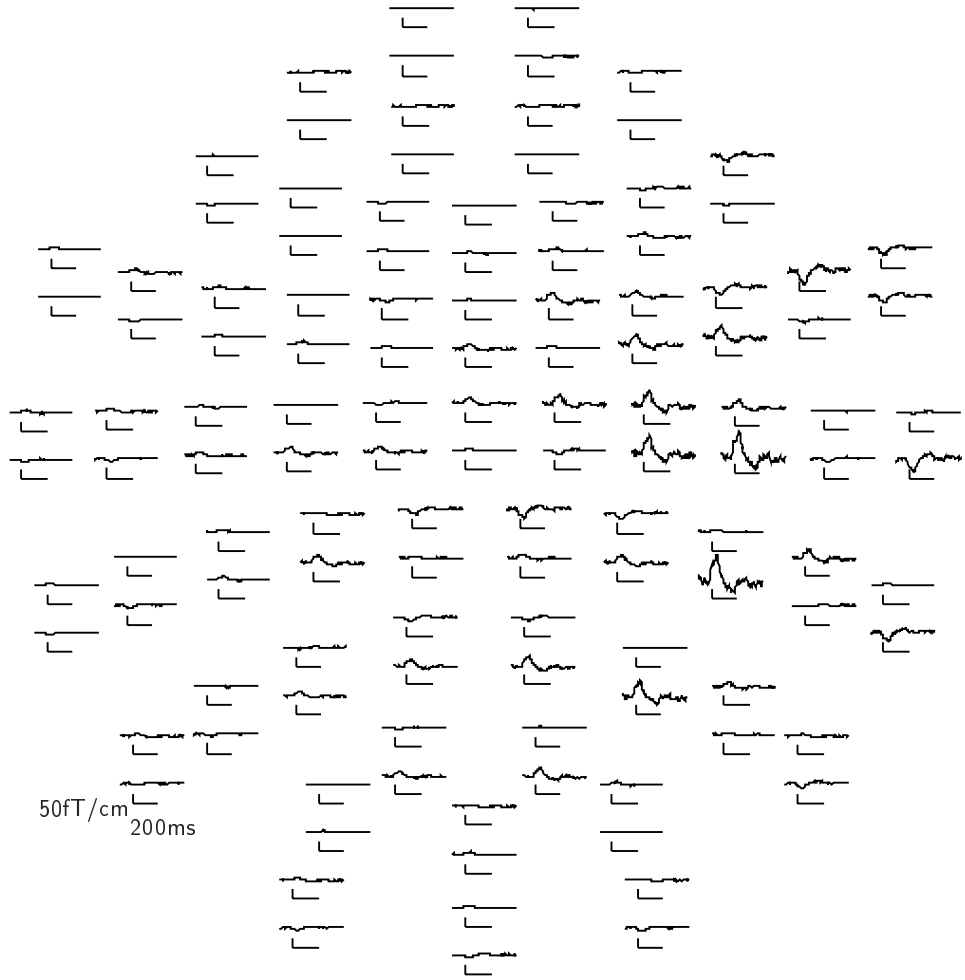


Figure 7.3 Sensor projection of component showing selective activation over the right fronto-parietal cortex ($N = 90$ trials, button press triggered averages). Compare with unseparated data in figure 7.5.

and the total energy in the sensor jump sources was 1.72±0.28% ($N = 13$). The ocular artifact and 60 Hz components have the most energy, while the energy in the neuronal components represented 10% or less of the total.

7.3.3 Localization of Separated Components: Examples

Using the sensor projection of task-related components as input to standard Neuromag dipole fitting software (xfit), we localized separated components. In conventional source localization practice, very often only 20–30 channels are selected for source localization. To show how well SOBI can isolate one neuronal source from another without relying on channel exclusion, throughout section 7.3.3–7.3.4, we

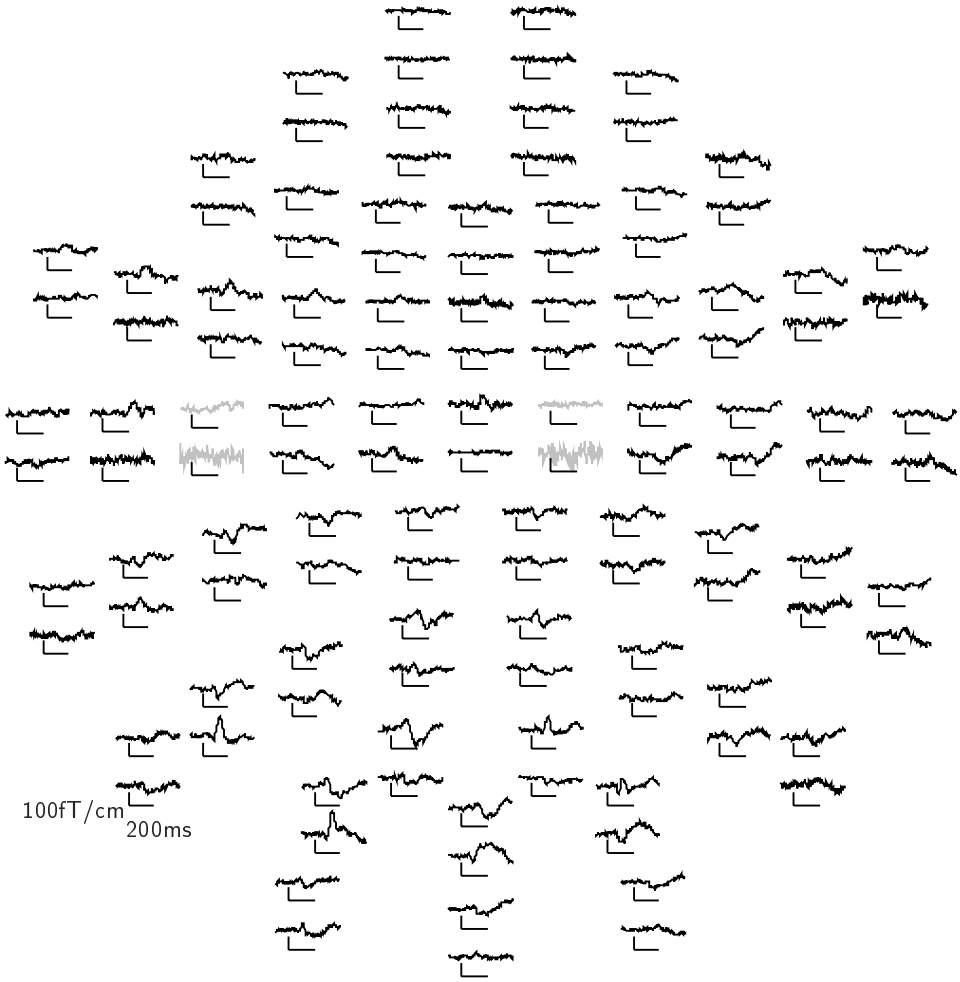


Figure 7.4 Visual stimulus triggered averages of unseparated data, $N = 90$ trials. Aberrant sensors are shaded.

Table 7.1 Range of energy accounted for (% of total energy across all channels) by the five categories of components

Category	Minimum	Maximum
Visual	0.21	24
Somatosensory	0.47	14
Ocular Artifact	0.57	72
60 Hz	0.26	44
Sensor Jump	0.17	12

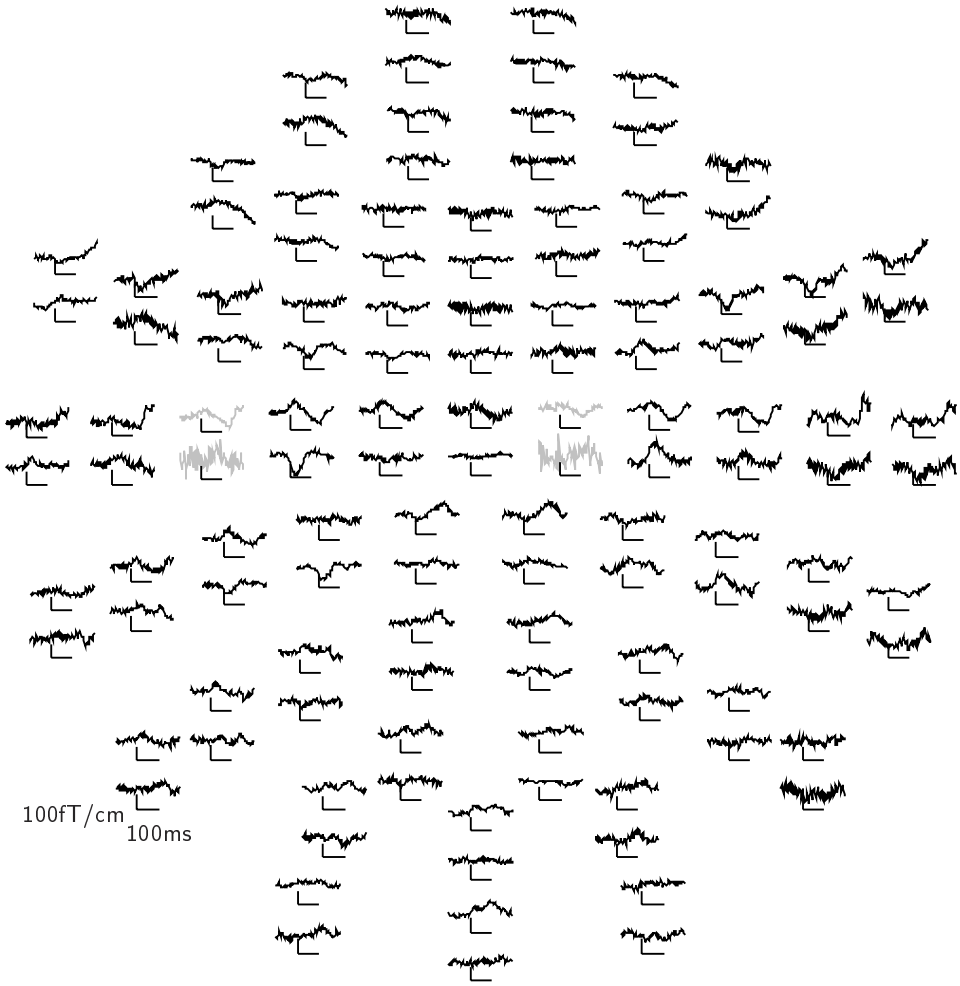


Figure 7.5 Button press triggered averages of unseparated data, $N = 90$ trials. Aberrant sensors are shaded.

generated the field maps, contour plots, and dipole localizations for components using *all* channels, *i.e.* without channel exclusion. To make the localization results comparable between using SOBI and without using SOBI, a subset of 20–30 channels were selected in dipole fitting in section 7.3.5, which provides statistical comparisons.

7.3.3.1 Visual Component

As the tasks involved simultaneous bilateral visual stimulation and judgment of its spatial location and identity, we expected SOBI to isolate visual components in the occipital, parietal, and temporal lobes. Components with visual evoked responses indeed showed field map activation over occipital, parietal and temporal lobes (not

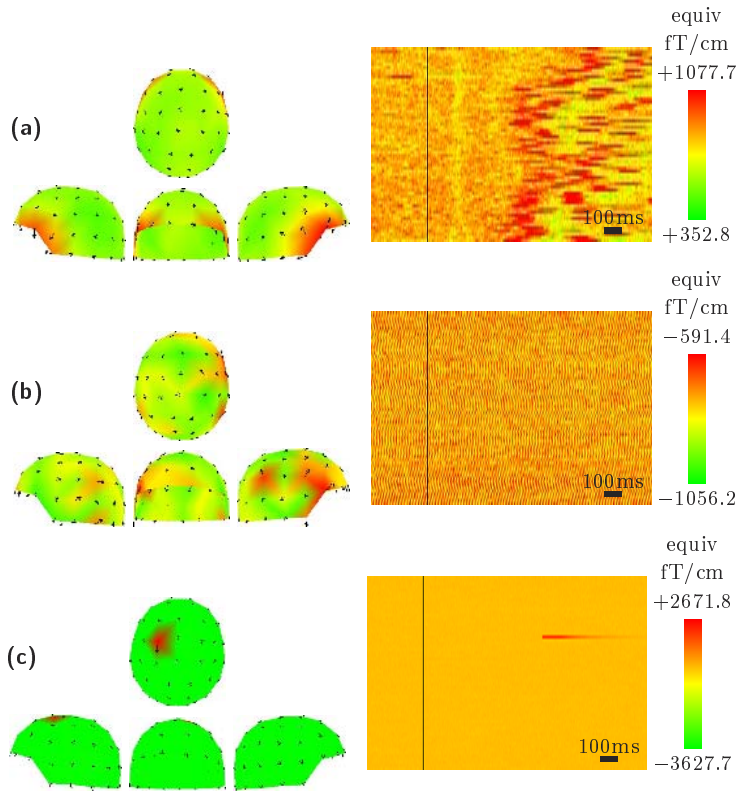


Figure 7.6 Field maps and unfiltered MEG images for **(a)** an ocular artifact component, **(b)** 60 Hz component, and **(c)** sensor jump component.

shown). For the particular stimuli used in these experiments, temporal sources were more variable in their precise location and temporal profiles. In contrast, occipito-parietal lobe activation appeared to have the greatest signal amplitudes and were reliably identified across multiple subjects. figure 7.7 left shows the dipole location of one such occipito-parietal visual source along with its field map and time course.

7.3.3.2 Somatosensory Component

As the tasks involved button presses, we expected both somatosensory and motor responses from the sensory and motor areas. figure 7.7 middle shows the dipole location of one component in the left hemisphere. Notice that this dipole is near the region where one finds dipoles from median nerve stimulation (Hari and Forss, 1999; Tesche and Karhu, 1997), and that the median nerve services the thumb. The time course of the response suggests that this response is unlikely to be a response from the motor cortex because the activations associated with motor preparation are typically estimated to be 385±85 ms before the movement onset (Hoshiyama et al., 1997), much earlier than the latency shown here. The

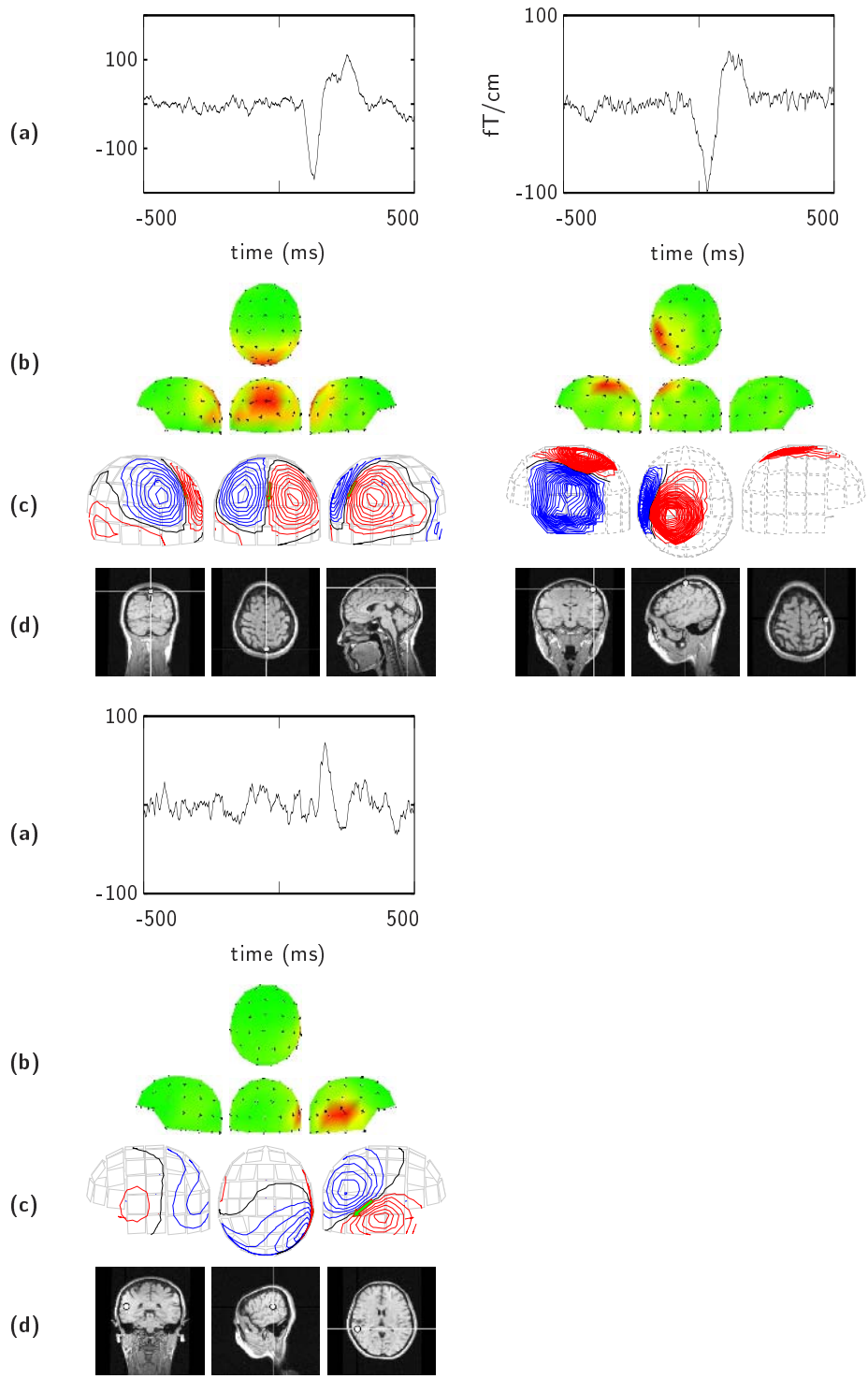


Figure 7.7 Examples of separated visual **(top left)**, somatosensory **(top right)**, and auditory **(bottom)** components, shown in **(a)** event triggered averages ($N = 90$ trials, stimulus onset at $t = 0$), **(b)** field maps, **(c)** contour plot, and **(d)** the fitted dipole superimposed on the subject's structural MRI images. All sensors (channels) were used in generating the contour plots and fitting the dipoles.

motor-evoked sensory responses with an estimated onset time of approximately 20±30 ms after the onset of movement (Hoshiyama et al., 1997) matched best to our button-press-elicited responses. Therefore, this component corresponds to a somatosensory source instead of a motor source. In contrast to typical fast-rising somatosensory responses recorded using median nerve stimulation (Hari and Forss, 1999; Tesche and Karhu, 1997), the slow-rising somatosensory responses recorded here were elicited by stimulation to the thumb due to button presses. These temporal profiles are expected to differ, due to the difference between a very brief and focal electric shock and a much longer and distributed stimulation to the thumb and its surrounding areas.

7.3.3.3 Auditory Component

As the tones were presented as feedback, auditory responses were expected from the auditory cortex. In contrast to the typical auditory responses recorded during a simple auditory oddball task, the auditory responses from our experiment were most likely to overlap with and perhaps to be affected by both visual, motor, and somatosensory processing. As both auditory responses and somatosensory responses were triggered on the button press, auditory responses needed to be distinguished from the somatosensory sources. The spatial location of their fitted ECDs in the auditory cortex and their longer response latencies were sufficient to allow the disambiguation. figure 7.7 right shows one unilateral auditory source, with slow-rising and long response latency, localized to the vicinity of the lateral fissure, as expected for auditory activation (Cansino et al., 1994).

The relatively longer response latency (∼180ms) may be due to particular aspects of the task (see section 7.5). Specifically, in order to process the auditory feedback the subjects must first switch their attention from the visual to the auditory modality, and this takes time. Furthermore, the subjects must process and further interpret the auditory stimulus in evaluating their behavioral response and registering the correct target stimuli into memory. This additional processing may account for the difference in the temporal profile of the auditory responses. As the tones were bilaterally presented, one would expect auditory components with field maps showing bilateral activation. The auditory components recovered in these experiments, however, were unilateral for two subjects and bilateral for the other two. This variability across subjects could be due to differences in cerebral dominance of auditory processing. A difference in the temporal aspect of the left and right auditory processing could be expected to lead to the identification of two separate left and right components.

In comparison to visual and somatosensory components, auditory components were much more difficult to identify, perhaps due to the above described complexity and associated variability. Although in most cases auditory components could be identified from visual inspection of the field map and event-triggered averages, the signal-to-noise ratios were too poor to permit consistent dipole fitting across tasks

and across subjects. Therefore, the following more detailed and systematic analysis of localization results will focus on only visual and somatosensory sources.

7.3.4 Cross-Task and Cross-Subject Reproducibility in Localization of Components

To show how reproducible the localization of components can be *across the four cognitive tasks,* we examined separated visual components from one subject. Across tasks, two occipito-parietal visual sources were reliably localized within the same subject from two separated components. For both visual sources, the time course of the response is highly repeatable across multiple tasks, as shown in the overlay plot (figure 7.8a,b). The earlier visual responses were almost identical in both amplitude and response latency (figure 7.8a), while the later responses varied only in amplitude across tasks (figure 7.8b). Given the number of subjects in this study (four), we do not have the statistical power to draw any conclusions about whether the amplitude increases monotonically with the complexity of the task.

These visual components identified from different tasks were localized to similar locations within the occipital and parietal lobes, as shown in figure 7.8c,d, in which fitted dipoles from multiple experiments are superimposed on the subject's structural MRI images. Notice that in the field map, the right side of the head is shown on the right whereas in the structural MRI images, following radiological convention, the right side is shown on the left.

To show how reproducible the localization of components can be *across subjects,* we examined separated somatosensory components from three subjects.[11] In all three subjects, we reliably identified two components (left and right) with button press locked responses in the somatosensory areas. figure 7.9 shows the time course, field map, contour plot, and fitted dipole for the somatosensory components in the right hemisphere of the three subjects. Notice the cross-subject similarity in the field maps, contour plots, and dipole locations (somatosensory cortex in the anterior parietal lobe, post-central sulcus).

7.3.5 Detecting Expected Neuronal Sources with and without SOBI

To offer quantitative comparison in the relative performance of source localization with and without SOBI, we attempted to identify and localize the most reliable occipito-parietal visual source, and both the left and right somatosensory sources in all subjects and all tasks from separated components and from the unprocessed data. As all four tasks involved bilateral presentation of visual stimuli, we expected that at least one visual source would be found active in the occipito-parietal cortex. Similarly, because separate left and right button presses were required by

11. The fourth subject did right-hand index-mid finger button presses which differed from the rest of the subjects.

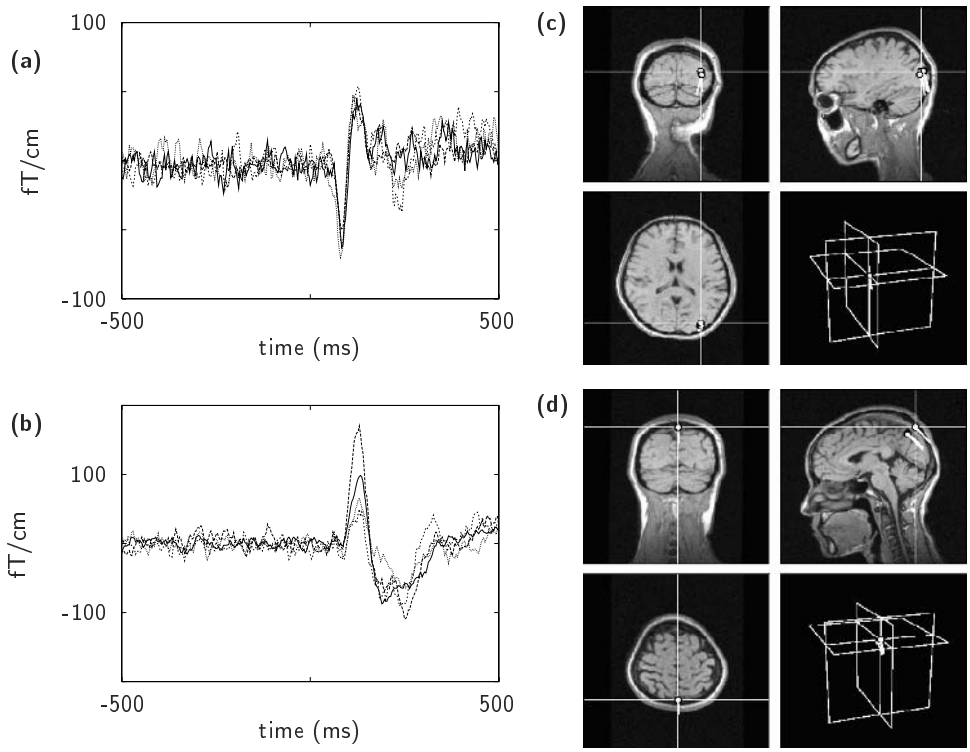


Figure 7.8 Cross-task consistency in the temporal profile **(a,b)** and dipole location **(c,d)** of two visual components. Occipital **(a,c)** and occipito-parietal **(b,d)** sources can be identified and localized consistently across multiple tasks (overlay). **(a,b)** Visual stimulus-triggered averages from 4 visual tasks, overlayed ($N = 90$ trials per task). **(c,d)** Corresponding single ECDs for visual sources in **(a,b)**. Notice consistency of the dipole locations across-tasks. Notice also the temporal profile of the earlier visual source **(a)** did not differ across tasks, but the amplitude of the later visual source **(c)** was modulated by the task conditions.

all the tasks, we also expected that at least one left and one right somatosensory source would be active. For these expected sources, we attempted to localize the source with dipole fitting from separated components and from the raw sensor data (without SOBI). The percentage of the expected sources for which dipole solutions can be found are compared for localization with and without the aid of SOBI.

For a component to be considered a detectable neuronal source, there must be an evoked response that clearly deviates from the baseline in the averaged component data. We rejected all components with any ambiguity on this criterion. Secondly, the components must have a field map showing focal activation of sensors over the relevant brain regions (occipito-parietal cortex and anterior parietal cortex in this study). Thirdly, the contour plot for the component must be dipolar. Finally, the

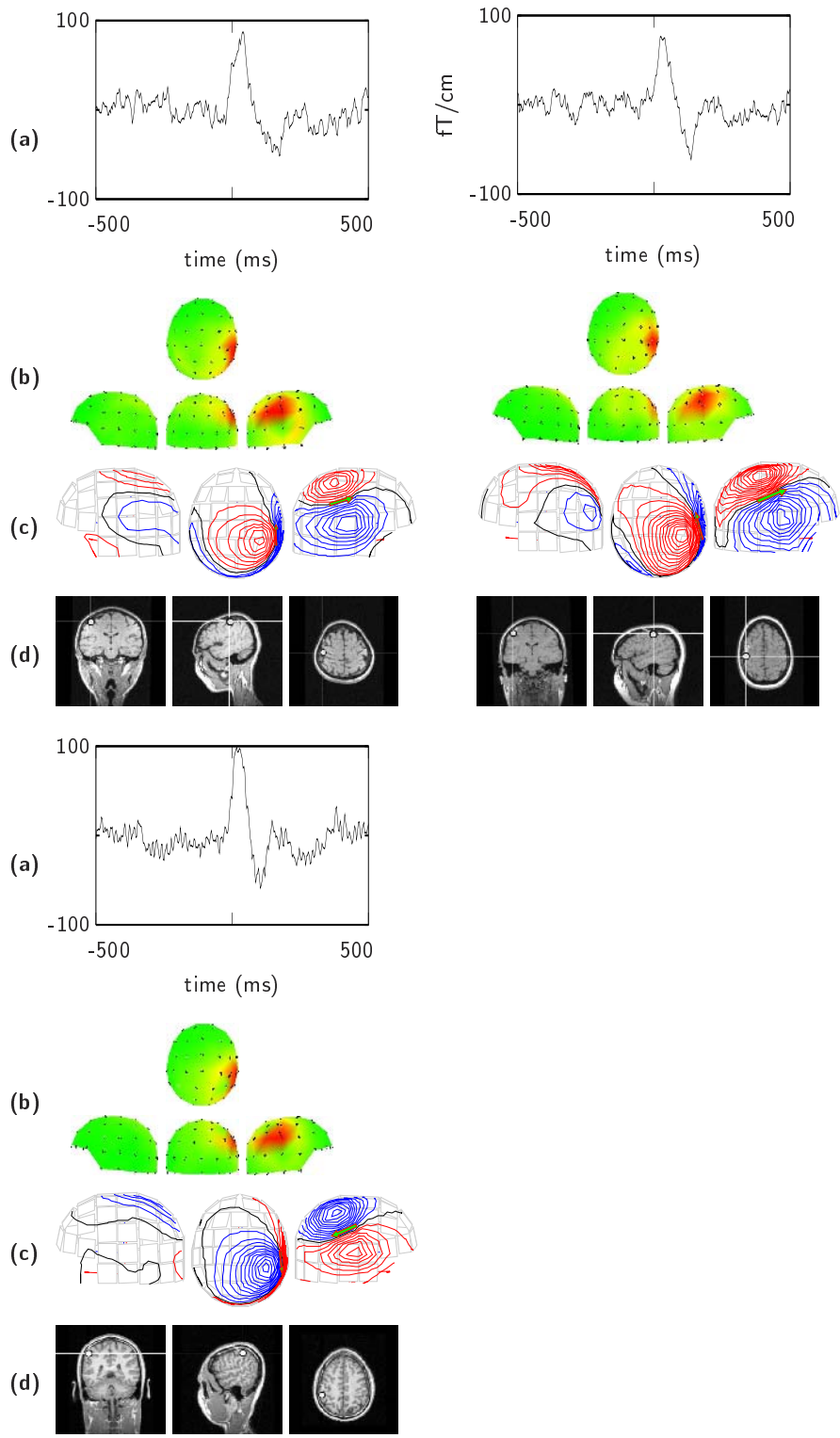


Figure 7.9 Somatosensory sources can be identified and localized consistently across multiple subjects. Similar to figure 7.7 except the responses were triggered by the button press.

fitted dipole must be in the relevant cortical areas. For a source to be considered detectable using the conventional method of localization, one must first identify a sensor at which the largest evoked response is found. Secondly, the contour plot must be dipolar at the peak time. Finally, in a few cases when multiple dipole solutions are needed, at least one of the dipoles is localized to the expected brain region.

7.3.5.1 Visual Sources

Among all separated components, for each subject and each task, we were able to identify and localize an occipito-parietal visual source with a single dipole (100% detectability). These occipito-parietal components invariantly had very focal sensor projections (see field maps in figure 7.7b), and the contour plots were invariably dipolar even without channel selection (for example, see field map and contour plot in figure 7.7a). Single dipoles were fitted for these occipito-parietal components. A subset of channels over the occipito-parietal lobe (20–30 channels) were used for the purpose of fair comparison with the conventional analysis method without the aid of SOBI. The peak response latencies of these components ($N = 16$) were 139.0 ± 7.6 and the dipole coordinates (X,Y,Z) were 7.5 ± 2.6, -49.4 ± 3.2, and 68.6 ± 3.4 mm.

Using the conventional method of source localization directly from the unseparated sensor data, dipoles were fitted using the same or similar subset of channels selected over the occipito-parietal cortex. In all subjects and all tasks, the conventional method identified and localized at least one visual source in the occipito-parietal lobe (100% detectability). Of a total of 16 expected sources (4 tasks by 4 subjects by 2 sides), 10 could be fitted with a single dipole, 4 were fitted with two-dipole solutions, 1 was fitted with a three-dipole solution, and 1 was fitted with a four-dipole solution. When multiple dipole solutions were needed, at least one of them was localized to the occipito-parietal cortex. This variation in dipole solutions may reflect some individual differences in visual processing occurring outside of the occipito-parietal cortex. The peak response latencies of these occipito-parietal visual sources ($N = 16$) were 143.6 ± 5.5 and the dipole coordinates (X,Y,Z) were 4.21 ± 4.8, -55.89 ± 2.68, and 59.42 ± 3.83 mm.

7.3.5.2 Somatosensory Source

From components of all subjects and all tasks, with only two failures we were able to identify and localize 22 out of the 24 expected left and right somatosensory sources with a single dipole (3 subjects by 4 tasks). All 22 somatosensory components had very focal sensor projections (see field maps in figure 7.9b) and their contour plots were all highly dipolar even without channel selection (for example, see field map and contour plot in figure 7.9.) Single dipoles were fitted to these components, with a subset of channels over the somatosensory cortex (20–30 channels) selected for the purpose of fair comparison with the conventional analysis method. The peak response latencies were 3.3 ± 4.2 and 0.8 ± 3.4 ms for the left ($N = 11$) and right

($N = 11$) somatosensory sources, respectively. The dipole coordinates (X,Y,Z) were -39.4 ± 2.4, 7.8 ± 2.7, and 84.6 ± 1.7 for the left and 45.69 ± 2.1, 5.6 ± 2.2, and 84.1 ± 3.1 for the right somatosensory sources.

Using the conventional method of source localization directly from the unseparated sensor data, dipoles were fitted using the same or similar subset of channels selected over the somatosensory cortex. Only 9 out of 24 expected left and right somatosensory sources could be identified and localized following the conventional method of identifying a peak response in the averaged sensor data. Of 24 sources expected (3 subjects by 4 tasks by 2 sides), in 7 cases no visible peak response could be identified in any of the sensors. Of the remaining 17 cases in which peak responses could be found in at least one sensor over the somatosensory cortex, 4 did not have dipolar fields, and 4 resulted in dipole locations outside of the head or in the auditory cortex. Single dipole solutions were found in only 9 cases. The peak response latencies of these somatosensory sources were -5.2 ± 2.5 for the left hemisphere ($N = 5$) and 1.6 ± 1.8 for the right hemisphere ($N = 4$). The dipole coordinates (X,Y,Z) were -43.3 ± 3.9, 12.1 ± 5.6, and 82.8 ± 3.8 for the left 42.3 ± 5.5, 15.9 ± 2.7 and 89.9 ± 1.4 for the right sources.

7.3.6 Statistical Comparisons

There was no significant difference in the detectability for the occipito-parietal source measured with and without SOBI. In contrast, SOBI preprocessing resulted in an increase in the detectability of the expected somatosensory sources (Chi Square, test $p < .0001$) (figure 7.10). The peak response latencies for the visual and somatosensory sources did not differ significantly when measured using and without using SOBI. For the visual sources, the precise dipole locations estimated with and without SOBI did not differ in the X and Y dimensions but nearly differed significantly in the Z dimension ($p = 0.05$). For the somatosensory sources, the precise dipole locations differed significantly in the Y dimension ($p < 0.05$) for the left source and in Y and Z dimensions for the right source ($p < 0.05$). As the true accuracy of source locations cannot be determined from these experiments without a depth-electrode, no quantitative comparisons can be made concerning accuracy.

7.4 Discussion

We identified and localized visual and somatosensory sources activated in four subjects during four cognitive tasks. Due to the relatively large variability involved in highly cognitive tasks and the small number of trials collected, these tasks were characterized by relatively poor signal-to-noise ratios in the sensor data and therefore were ideal for evaluating differential localization performance. Our results showed that despite the large variability associated with the visual and somatosensory activations during these particular tasks, SOBI was able to separate identifiable visual and somatosensory components that were further localized to the

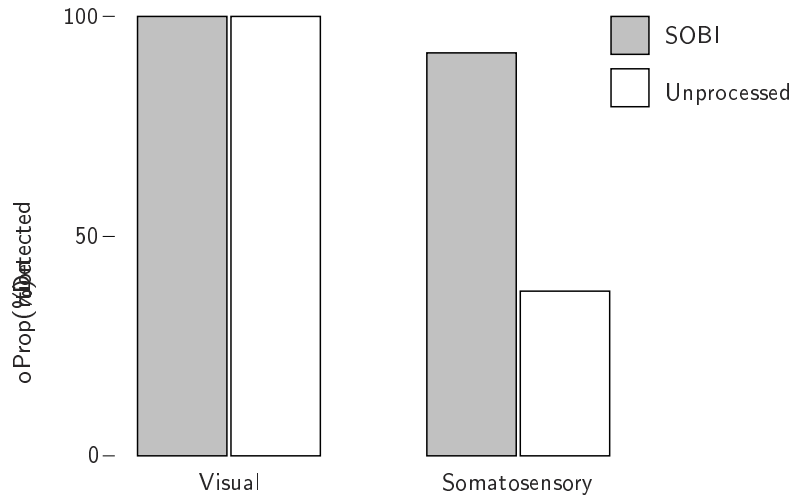


Figure 7.10 SOBI increased the detectability of expected neuronal sources for the more variable somatosensory activation.

expected cortical regions. The physiological and neuroanatomical interpretability of these components across multiple sensory modalities and their cross-subject and cross-task reproducibility establish SOBI as a viable method for separating and identifying neuronal populations from MEG data during fairly complex cognitive tasks. Most importantly, we showed that SOBI preprocessing offered a special advantage when the evoked responses in the sensor data had poor signal-to-noise ratios. Specifically, for the highly variable somatosensory activation evoked by incidental stimulation during button presses, SOBI preprocessing resulted in a greater percentage of the expected somatosensory sources being identified and localized than the same dipole modeling method applied directly to the raw sensor data.

7.4.1 SOBI Reduced Subjectivity and Labor in Source Localization

In conventional source localization, there are two major sources of subjectivity: the selection of dipole fitting times, and the selection of channels. These are both eliminated by our proposed procedure. First, because each component has a fixed field map, the dipole fitting solutions for components were not sensitive to either the time at which the dipoles were fitted nor to the sensor used for determining the time of fit (see section 7.6.2). Within this map, each sensor reading reflects only activation due to a single source generator, or several temporally coherent generators as opposed to activation due to a combination of multiple generators, each with a different time course. Therefore, using SOBI, there is no need to subjectively select a time from a sensor for dipole fitting. Secondly, simple components, which have field activation over early sensory processing areas, were almost always dipolar even

without channel selection/reduction.[12] Therefore, channel selection is not necessary. One way to see the difference between dipole localization with and without SOBI processing is to view SOBI as a more automatic and more objective tool that allows the isolation of sensor activation due to an already isolated functionally independent generator. The reduced subjectivity and time required to find dipole solutions can make data analysis and training of new researchers for MEG more cost-effective.

7.4.2 SOBI Improved Detectability of Neuronal Sources

The advantages of ICA algorithms in general have been shown in a number of applications to EEG and MEG data. First, these algorithms can separate neuronal activity from various artifacts (Makeig et al., 1996; Vigário et al., 1998; Tang et al., 2000a; Ziehe et al., 2000; Jung et al., 2000a,b), such as eye blinks. In contrast to methods that rely on the use of a template, ICA removes these artifacts without any prior assumptions about the nature of the waveforms. Secondly, ICA isolates physiologically and behaviorally meaningful components that describe previously unavailable aspects of neuronal activity (Makeig et al., 1997, 1999b; Wübbeler et al., 2000). Finally, ICA-separated neuronal sources are less contaminated by various noise sources, which allows single-trial response detection (Jung et al., 1999; Tang et al., 2000b; Carter et al., 2000). ICA methods have been able to distinguish the absence of rhythmic activity from the absence of phase locked rhythmic activity (Makeig et al., 1999a).

We have shown that SOBI separation of the data resulted in a greater detectability of somatosensory sources, but did not increase the detectability of visual sources. This modality-specific improvement in source detectability depended on the signal-to-noise ratios in the sensor data. Because visual responses could be clearly identified from the raw sensor data even without the aid of SOBI, it would not have been possible for SOBI to improve the detection rate. In contrast, the relatively poor signal-to-noise ratios in the raw sensor data for the somatosensory responses caused many failures in identifying a sensor at which a peak response occurred and in determining the peak response time. Under this poor signal-to-noise condition, in all but two cases, SOBI preprocessing resulted in separated components with the characteristic field map, characteristic temporal response profile, and the correct dipole location for a somatosensory source. These findings suggest another advantage that ICA algorithms can offer: improving the ability to detect and localize neuronal sources that are otherwise difficult to detect or are undetectable under relatively poor signal-to-noise conditions.

This improvement has significant practical implications. First, brain regions involved in higher level cognitive processing tend to show greater trial-to-trial

12. SOBI also separated out many complex components which have multiple patches or very broad field activation. These components may reflect synchronized activation in multiple brain regions. Functional connectivity may be inferred among these brain regions.

variability in their activation, and therefore, have lower signal-to-noise ratios in the average responses. Second, behavioral tasks that bear greater resemblance to real world situations tend to involve greater variability in both stimulus presentation and subsequent processing. Finally, studies of clinical patients and children are often limited by the length of the experiment, and therefore, often provide data from a limited number of trials. Our results suggest that ICA may offer an improved capability in detecting and localizing neuronal source activations in these difficult situations.

7.4.3 Assumptions of SOBI

Here, we discuss assumptions of particular relevance to SOBI and MEG, rather than general issues in ICA. Like all ICA algorithms, SOBI assumes that the mixing process is stable. In the context of MEG, a stable mixing process corresponds to assuming that the head is motionless relative to the sensors. For this reason head stabilization can be particularly important in MEG when ICA is used. SOBI also assumes that there are at least as many sensors as sources. For us, this is not a serious problem, as our MEG device has 122 sensors, yet we recover only a few dozen sources that show task-related evoked responses. The observation that only a small number of sources are active during typical cognitive and sensory activation tasks is consistent with the results of studies using both EEG (Makeig et al., 1999b) and MEG (Vigário et al., 2000). The crucial assumption in ICA is that of independence. For a thorough discussion of the independence assumption as it pertains to MEG, see Vigário et al. (2000). Here, we will discuss independence only in the context of the particular measure of independence used by SOBI.

One problem that EEG and MEG researchers have with the independence assumption arises from the fact that if one computes correlations between EEG or MEG sensor readings over multiple brain regions during behavioral tasks, one would find that some brain regions have non-zero correlations. A good example of correlated brain activity is the apparently correlated evoked responses from neuronal populations in multiple visual areas along the processing pathway during a visual stimulus presentation. Based on such an observation, one could conclude that as the statistical independence assumed by ICA is clearly violated, the results of ICA must not be trusted. Yet, we have shown that SOBI was able to separate visual components that clearly correspond to neuronal responses from early and later visual processing stages that are correlated due to common input (Tang et al., 2000b). Others (Makeig et al., 1999b; Vigário et al., 2000) have produced behaviorally and neurophysiologically meaningful components under a variety of task conditions.

As different ICA algorithms use the independence assumption differently, we offer the following explanation that applies specifically to SOBI. One needs to recognize that correlation is not a binary quantity. Consequently, neither is violation of the independence assumption. The important question is not whether the assumption is violated but whether the assumption is *sufficiently* violated such that the estimated

neuronal sources by SOBI are no longer meaningful. The way SOBI uses the independence assumption is to minimize the total correlations computed with a set of time delays, as described in section 7.6.1. As such, each delay-correlation matrix $\mathbf{R}_\tau$ generally makes only a small contribution to the objective function. For example, the correlation one would observe between V1 and V2 responses could be high only at or around one particular time delay, say in $\mathbf{R}_{20\text{ms}}$. In optimizing its objective function, SOBI can leave a particularly large non-zero off-diagonal element, say the one corresponding to the 20ms delayed correlation between V1 and V2, when minimizing the sum squared off-diagonal elements across all the components and time delays. Therefore, this particular method of maximizing independence is not necessarily incompatible with a large correlation at a particular time delay between two sources sharing common inputs.

Most ICA algorithms, including SOBI, minimize some objective function. It is possible for the optimization process to find a poor local minimum. In general, poor results can result from many underlying causes: poor experimental design, poorly conducted experiments, poor head stabilization, poor optimization within the ICA algorithm, violation of assumptions, etc. No amount of attention to any one possible problem can validate ICA-based methods for processing functional brain imaging data. As with any statistical procedure, the real issue here should not be whether assumptions are violated at all, but whether the algorithms can robustly produce separated components that are behaviorally, neuroanatomically, and physiologically interpretable, despite some violation of the assumptions under which the algorithms were derived. For example, t-tests are very robust against the violation of normality assumption and are therefore regularly performed on data which are not guaranteed to be Gaussian. Only empirical results can give confidence that a method is correctly separating the MEG data.

7.4.4 Summary

Establishing that (1) SOBI preprocessing can lead to the identification and localization of physiologically and anatomically meaningful neuronal sources and (2) SOBI preprocessing can increase the success rate in detecting and localizing neuronal source activation under poor signal-to-noise conditions is only the first step in demonstrating the usefulness of ICA algorithms to the analysis and interpretation of MEG data. The next steps include systematically studying the effect of ICA on source localization when ICA methods are combined with more sophisticated source localization algorithms (Ribary et al., 1991; Aine et al., 1998; Mosher and Leahy, 1999; Schmidt et al., 1999) and exploring the possibility of measuring single-trial response onset times in ICA separated neuronal sources.

Acknowledgments

This Chapter is reprinted with permission of The MIT Press from *Neural Computation* 14, no. 8 (Aug. 2002): 1827 $\simeq$ 1858.

Supported by the National Foundation for Functional Brain Imaging and NSF CAREER award 97-02-311, an equipment grant from Intel corporation, the Albuquerque High Performance Computing Center, a gift from George Cowan, and a gift from the NEC Research Institute. We thank Ole Jensen for tips on packaging the separated components for the Neuromag software, Mike Weisend for granting us access to his data, and Robert Christner for technical support. We also thank Lloyd Kaufman, Zhonglin Liu, Claudia Tesche, Cheryl Aine, and Roland Lee for their comments and discussion.

7.5 Experimental Details

Continuous 122-channel data were collected during the entire period of the following four tasks sampled at 300 Hz and band-pass filtered at 0.03–100 Hz. A total of four visual reaction time tasks were performed by each subject.

In all tasks, each trial consisted of a pair of colored abstract block compositions, one of which was the target, presented symmetrically and simultaneously on the left and right halves of the screen. Subjects were instructed to respond as quickly and as accurately as possible with a left or right hand mouse button press when the target stimulus was presented to the left or right side of the display screen respectively. The button press elicited an auditory feedback indicating whether a correct or incorrect response was made.

Stimuli were either presented on 15 inch VGA computer monitor at a distance of 48 inch and occupying 7.6° of visual angle or back-projected by an LCD projector positioned so that the stimulus occupied the same visual angle. In all tasks the interval between the motor-response and the next stimulus presentation was 3.0±0.5s. Auditory feedback was composed of 2000 Hz and 500 Hz tones indicating correct and incorrect choices respectively.

The four tasks differed from each other primarily in their definition of the target stimulus which affected how much processing was required for target determination. The precise duration of each task varied slightly across subjects, depending upon the subject's reaction time. Therefore typical durations are given below. The first task (stimulus pre-exposure) consisted of 270 trials. It took the subjects approximately 30 minutes to perform this task. The other three tasks (elemental discrimination, trump card, and transverse patterning) each consisted of 90 trials that were subsets of the same stimuli contained in the first task. Each of the these three tasks took approximately 10 minutes to complete. For each subject, all four experiments were performed on the same day but each in a separate session. Instructions for each experiment were given immediately prior to that experiment. Subjects were permitted to move between experiments. Head positions were recalibrated at the beginning of each experiment. Subjects performed the four experiments in order of increasing task demand: *stimulus pre-exposure*, *trump-card task*, *elemental discrimination task*, and *transverse patterning task*.

7.5.1 Stimulus Pre-Exposure Task

There were no pre-defined relationships between stimuli and button presses. No feedback was given to the subjects about any choice. The subject was instructed to examine both stimuli and then make a roughly equal number of right and left button presses, without consistent alternation between right and left responses. The sequence of presentation was random. Presentations of each stimuli on the left and right sides of the video screen were counterbalanced.

7.5.2 Trump Card Task

Subjects were instructed to discover by trial and error which of the two stimuli in the stimulus pair was the target (the trump card). A total of 9 stimulus pairs involving 10 stimuli were used, with a single stimulus as the trump card. Subjects did not have any problem in discovering the trump-card within a few trials.

7.5.3 Elemental Discrimination Task

Subjects were instructed to discover which one of the stimulus pair was the target stimulus by trial and error. A total of three stimulus pairs consisting of six stimuli were used. For each pair of stimuli, one of the pair was the target. This task differs from the trump card task in that multiple target stimuli were involved. All subjects found the targets within a few trials.

7.5.4 Transverse Patterning Task

Subjects were instructed to discover which of the two stimuli in a stimulus pair was the target. Three stimulus pairs consisting of three stimulus compositions were used. Each stimulus could be a target or non-target depending upon what it was paired with. The target definition was a "rock-paper-scissors" arrangement: A wins when paired with B, B wins when paired with C, C wins when paired with A. Again, subjects were able to discover the winning relationships after a few trials.

7.6 Mathematical Methods

7.6.1 The SOBI Source Separation Algorithm

The SOBI algorithm (Belouchrani et al., 1993) proceeds in two stages. First, the sensor signals are zero-meaned and pre-sphered as follows:

$$\mathbf{y}(t) = \mathbf{B}\left(\mathbf{x}(t) - \langle \mathbf{x}(t) \rangle\right) \tag{7.2}$$

The angle brackets $\langle \cdot \rangle$ denote an average over time, so the subtraction guarantees that $\mathbf{y}$ will have a mean of zero. The matrix $\mathbf{B}$ is chosen so that the correlation

matrix of $\mathbf{y}$, namely $\langle \mathbf{y}(t)\, \mathbf{y}(t)^T \rangle$, becomes the identity matrix. This is accomplished by moving to the PCA basis using $\mathbf{B} = \mathrm{diag}(\lambda_i^{-1/2})\mathbf{U}^T$, where λ_i are the eigenvalues of the correlation matrix $\langle (\mathbf{x}(t) - \langle \mathbf{x}(t) \rangle)\, (\mathbf{x}(t) - \langle \mathbf{x}(t) \rangle)^T \rangle$ and $\mathbf{U}$ is the matrix whose columns are the corresponding eigenvectors, *i.e.* the "PCA components" of $\mathbf{x}$. (This pre-sphering is solely for the purpose of improving the numerics of the situation by constraining the matrix $\mathbf{V}$ below to be a rigid rotation.)

For the second stage, one constructs a set of matrices which, in the correct separated basis, should be diagonal. In our case, we chose a set of time-delay values τ to compute symmetrized correlation matrices between the signal $\mathbf{y}(t)$ and a temporally shifted version of itself,

$$\mathbf{R}_\tau = \mathrm{sym}(\langle \mathbf{y}(t)\, \mathbf{y}(t+\tau)^T \rangle) \tag{7.3}$$

where $\mathrm{sym}(\mathbf{M}) = (\mathbf{M} + \mathbf{M}^T)/2$ is a function that takes an asymmetric matrix and returns a closely related symmetric one. This symmetrization discards some information, but the problem is already highly over-constrained, and the symmetrized matrices provide valid, albeit slightly weaker, constraints on the solution.

After calculating the $\mathbf{R}_\tau$, we look for a rotation $\mathbf{V}$ that jointly diagonalizes all of them by minimizing $\sum_\tau \sum_{i \neq j} (\mathbf{V}^T \mathbf{R}_\tau \mathbf{V})_{ij}^2$, the sum of the squares of the off-diagonal entries of the matrix products $\mathbf{V}^T \mathbf{R}_\tau \mathbf{V}$, via an iterative process (Cardoso and Souloumiac, 1996, using MATLAB code available at http://sig.enst.fr/~cardoso/). The final estimate of the separation matrix is $\mathbf{W} = \mathbf{V}^T\, \mathbf{B}$, which is used to calculate the separated components $\hat{\mathbf{s}}(t) = \mathbf{W}\, \mathbf{x}(t)$.

7.6.2 Separated Components in Sensor Space

Since $\mathbf{W}$ is the estimated unmixing matrix, let us use $\hat{\mathbf{s}}(t) = \mathbf{W}\, \mathbf{x}(t)$ for the consequent estimated sources, and $\hat{\mathbf{A}} = \mathbf{W}^{-1}$ for the corresponding estimated mixing matrix. Using these, the sensor signals resulting from just one of the components can be computed as $\hat{\mathbf{x}}(t) = \hat{\mathbf{A}}\, \mathbf{D}\, \mathbf{W}\, \mathbf{x}(t) = \hat{\mathbf{A}}\, \mathbf{D}\, \hat{\mathbf{s}}(t)$, where $\mathbf{D}$ is a matrix of zeros except for ones on the diagonal entries corresponding to each component which is to be retained.

To localize a single component, one computes

$$\hat{\mathbf{x}}^{(i)}(t) = \hat{s}_i(t)\, \hat{\mathbf{a}}^{(i)} \tag{7.4}$$

where $\hat{\mathbf{a}}^{(i)}$ is the i^{th}column of $\hat{\mathbf{A}}$ and $\hat{\mathbf{x}}^{(i)}(t)$ is the sensor-space image of source i. Because $\hat{\mathbf{x}}^{(i)}(t)$ is at each point in time equal to the unchanging vector $\hat{\mathbf{a}}^{(i)}$, scaled by the time course $\hat{s}_i(t)$, dipole fitting algorithms will localize $\hat{\mathbf{x}}^{(i)}(t)$ to the same location no matter what window in time is chosen.

7.6.3 Scaling

Blind source separation leaves the freedom to choose an arbitrary scale factor for each component. For instance, the source $s_i(t)$ could be scaled up by a factor of ten, and the i^{th} column of $\mathbf{A}$ scaled down by the same factor of ten, giving rise

to the exact same observation $\mathbf{x}(t)$. Making a reasonable assumption that all the sensors have intrinsic Gaussian noise of the same magnitude, we used the additivity of these independent sensor noises to scale each row of $\mathbf{W}$ to give each row a vector length of one. That is, if $\widetilde{\mathbf{W}}$ is the unscaled unmixing matrix, then we normalized its rows to yield $\mathbf{W}$ using

$$w_{ij} = \frac{\tilde{w}_{ij}}{\sqrt{\sum_j \tilde{w}_{ij}^2}} \tag{7.5}$$

With this scale factor, the sources can be viewed as being measured by a "virtual sensor" that measures in the same units, with the same scale, and with the same amount of intrinsic noise as the real sensors. This gives rise to "effective Ft/cm" units as above.

An alternative approach to scaling is to try to calculate the actual strength of the source, for instance the actual total energy emitted. This can be done by fitting a physical source model (such as an equivalent current dipole) to each component and scaling the rows of $\mathbf{W}$ such that the columns of $\hat{\mathbf{A}}$ match the attenuations predicted by the estimated physical model. This approach has the disadvantage of being dependent on the localization process, thus giving rise to multiple scalings when there are multiple localization procedures in use, or even when a single procedure produces multiple possible localizations. Another disadvantage of this alternative is a failure to generate a scaling when the localization fails, as it would on noise components.

7.6.4 Energy/Variance Accounted For

A commonly used statistic is the energy in a source, or the amount of variance it accounts for. The energy of source i is

$$E_i = \sum_t \sum_j (\hat{x}_j^{(i)}(t) - \overline{\hat{x}_j^{(i)}})^2 \tag{7.6}$$

where the mean is being subtracted to discount DC offsets, an important consideration in MEG. Because the rows of the matrix $\mathbf{W}$ are normalized, we can simplify this expression using Equation 7.4 yielding

$$E_i = \sum_t (\hat{s}_i(t) - \overline{\hat{s}_i})^2 \tag{7.7}$$

which is computationally more efficient. In this chapter, we gave the fraction of variance accounted by the i^{th}component as $E_i / \sum_i E_i$.

References

C. Aine, M. Huang, J. Stephen, and R. Christner. Multistart algorithms for MEG empirical data analysis reliably characterize locations and time courses of multiple sources. *Neuroimage*, 12(2):159–172, 2000.

C. J. Aine, M. Huang, R. Christner, J. Stephen, J. Meyer, J. Silveri, and M. Weisend. New developments in source localization algorithms: Clinical examples. *International Journal of Psychophysiology*, 30:198, September 1998.

S.-I. Amari and A. Cichocki. Adaptive blind signal processing—neural network approaches. *Proceedings of the IEEE*, 9, 1998.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

A. Belouchrani, K. A. Meraim, J.-F. Cardoso, and E. Moulines. Second-order blind separation of correlated sources. In *Proc. Int. Conf. on Digital Sig. Proc.*, pages 346–351, Cyprus, 1993.

S. Cansino, S. J. Williamson, and D. Karron. Tonotopic organization of human auditory association cortex. *Brain Res.*, 663:38–50, 1994.

J. T. Cao, N. Murata, S. Amari, A. Cichocki, T. Takeda, H. Endo, and N. Harada. Single-trial magnetoencephalographic data decomposition and localization based on independent component analysis approach. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E83A(9):1757–1766, 2000.

J.-F. Cardoso. On the performance of orthogonal source separation algorithms. In *European Signal Processing Conference*, pages 776–779, Edinburgh, September 1994.

J.-F. Cardoso. Blind signal separation: statistical principles. *Proceedings of the IEEE*, 9(10):2009–2025, October 1998. URL ftp://sig.enst.fr/pub/jfc/Papers/ProcIEEE.us.ps.gz.

J.-F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM Journal of Matrix Analysis and Applications*, 17(1), January 1996.

S. A. Carter, A. C. Tang, B. A. Pearlmutter, L. K. Anderson, C. J. Aine, and R. Christner. Co-activation of visual and auditory pathways induces changes in the timing of evoked responses in populations of neurons: an MEG study. *Society for Neuroscience Abstracts*, 26(9197), 2000.

J. J. Ermer, J. C. Mosher, M. X. Huang, and R. M. Leahy. Paired MEG data set source localization using recursively applied and projected (RAP) MUSIC. *IEEE Transactions on Biomedical Engineering*, 47(9):1248–1260, 2000.

J. S. George, C. J. Aine, J. C. Mosher, D. M. Schmidt, D. M. Ranken, H. A. Schlitt, C. C. Wood, J. D. Lewine, J. A. Sanders, and J. W. Belliveau. Mapping function in the human brain with magnetoencephalography, anatomical magnetic resonance imaging, and functional magnetic resonance imaging. *J. Clin. Neurophysiol.*, 12:406–431, 1995.

M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Modern Physics*, 65:413–497, 1993.

R. Hari and N. Forss. Magnetoencephalography in the study of human somatosensory cortical processing. *Phil. Trans. R. Soc. Lond. B*, 354:1145–1154, 1999.

R. Hari and R. Salmelin. Human cortical oscillations: a neuromagnetic view through the skull. *Trends in Neuroscience*, 20:44–49, 1997.

M. Hoshiyama, R. Kakigi, P. Berg, S. Koyama, Y. Kitamura, M. Shimojo, S. Watanabe, and A. Nakamura. Identification of motor and sensory brain activities during unilateral finger movement: Spatiotemporal source analysis of movement-associated magnetic fields. *Exp. Brain Res.*, 115(1):6–14, 1997.

M. X. Huang, C. Aine, L. Davis, J. Butman, R. Christner, M. Weisend, J. Stephen, J. Meyer, J. Silveri, M. Herman, and R. R. Lee. Sources on the anterior and posterior banks of the central sulcus identified from magnetic somatosensory evoked responses using multi-start spatio-temporal localization. *Human Brain Mapping*, 11(2):59–76, 2000.

A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999.

A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7), October 1997.

T.-P. Jung, C. Humphries, T.-W. Lee, M. J. McKeown, V. Iragui, S. Makeig, and T. J. Sejnowski. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37:163–178, 2000a.

T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, 111(10):1745–1758, 2000b.

T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and Terrence J. Sejnowski. Analyzing and visualizing single-trial event-related potentials. In *Advances in Neural Information Processing Systems 11*, pages 118–124. MIT Press, 1999.

Y. Kinouchi, G. Ohara, H. Nagashino, T. Soga, F. Shichijo, and K. Matsumoto. Dipole source localization of MEG by BP neural networks. *Brain Topogr.*, 8(3): 317–321, 1996.

J. D. Lewine and W. W. Orrison, II. Magnetoencephalography and magnetic source imaging. In W. W. Orrison, II, J. D. Lewine, J. A. Sanders, and M. F. Hartshorne, editors, *Functional Brain Imaging*, pages 369–417. Mosby, St. Louis, 1995.

S. Makeig, T. P. Jung, A. J. Bell, D. Ghahremani, and T. J. Sejnowski. Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci. USA*, 94(20):10979–10984, 1997.

S. Makeig, J. Townsend, T-P Jung, S. Enghoff, C. Gibson, T. J. Sejnwoski, and E. Courchesne. Early visual evoked response peaks apper to be sums of activity in multiple alpha sources. *Soc. Neurosci. Abstr.*, 25(652.8), 1999a.

S. Makeig, M. Westerfield, T-P Jung, J. Covington, J. Townsend, T. J. Sejnwoski, and E. Courchesne. Functionally independent components of the late positive event-related potential during visual spatial attention. *J. Neurosci.*, 19(7):2665–2680, 1999b.

S. Makeig, M. Westerfield, J. Townsend, T-P. Jung, E. Courchesne, and T. J. Sejnowski. Functionally independent components of the early event-related potential in a visual spatial attention task. *Philosophical Transactions of the Royal Society: Biological Sciences*, 354:1135–44, 1999c. URL ftp://ftp.cnl.salk.edu/pub/scott/N1paper.ps.gz.

S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski. Independent component analysis of electroencephalographic data. In *Advances in Neural Information Processing Systems 8*, pages 145–151. MIT Press, 1996. URL http://www.cs.cmu.edu/Groups/NIPS/NIPS95/Papers.html.

J. C. Mosher and R. M. Leahy. Recursive music: A framework for EEG and MEG source localization. *IEEE Transactions on Biomedical Engineering*, 45(11):1342–1354, 1998.

J. C. Mosher and R. M. Leahy. Source localization using recursively applied projected (RAP) MUSIC. *IEEE Transactions on Signal Processing*, 47(2):332–340, 1999.

J. C. Mosher, P. S. Lewis, and R. M. Leahy. Multiple dipole modeling and localization from spatiotemporal MEG data. *IEEE Transactions on Biomedical Engineering*, 39(6):541–557, 1992.

T. Nagano, Y. Ohno, N. Uesugi, H. Ikeda, A. Ishiyama, and N. Kasai. Multi-source localization by genetic algorithm using MEG. *IEEE Transactions on Magnetics*, 34(5/pt.1):2976–2979, 1998.

B. A. Pearlmutter and L. C. Parra. A context-sensitive generalization of ICA. In *International Conference on Neural Information Processing*, pages 151–157, Hong Kong, September 24–27 1996. Springer-Verlag. URL http://www.cs.unm.edu/~bap/papers/iconip-96-cica.ps.gz.

U. Ribary, A. A. Ioannides, K. D. Singh, R. Hasson, J. P. R. Bolton, F. Lado, A. Mogilner, and R. Llinas. Magnetic-field tomography of coherent thalamocortical 40-Hz oscillations in humans. *Proc. Natl. Acad. Sci. USA*, 88(24):11037–11041, 1991.

D. M. Schmidt, J. S. George, and C. C. Wood. Bayesian inference applied to the electromagnetic inverse problem. *Human Brain Mapping*, 7(3):195–212, 1999.

D. P. Schwartz, J. M. Badier, P. Bihoue, and A. Bouliou. Evaluation of a new MEG–EEG spatio-temporal localization approach using a realistic source model. *Brain Topogr.*, 11(4):279–289, 1999.

K. Sekihara, S. S. Nagarajan, D. Poeppel, S. Miyauchi, N. Fujimaki, H. Koizumi, and Y. Miyashita. Estimating neural sources from each time-frequency component of magnetoencephalographic data. *IEEE Transactions on Biomedical Engineering*, 47(5):642–653, 2000.

K. Sekihara, D. Poeppel, A. Marantz, H. Koizumi, and Y. Miyashita. Noise covariance incorporated MEG-MUSIC algorithm: A method for multiple-dipole estimation tolerant of the influence of background brain activity. *IEEE Transactions on Biomedical Engineering*, 44(9):839–847, 1997.

A. C. Tang, B. A. Pearlmutter, M. Zibulevsky, and S. A. Carter. Blind separation of multichannel neuromagnetic responses. *Neurocomputing*, 32–33:1115–1120, 2000a.

A. C. Tang, B. A. Pearlmutter, M. Zibulevsky, T. A. Hely, and M. P. Weisend. An MEG study of response latency and variability in the human visual system during a visual-motor integration task. In *Advances in Neural Information Processing Systems 12*, pages 185–191. MIT Press, 2000b.

C. D. Tesche and J. Karhu. Somatosensory evoked magnetic fields arising from sources in the human cerebellum. *Brain Research*, 744(1):23–31, 1997.

K. Uutela, M. Hämäläinen, and R. Salmelin. Global optimization in the localization of neuromagnetic sources. *IEEE Transactions on Biomedical Engineering*, 45(6): 716–723, 1998.

R. Vigário, J. Särelä, V. Jousmaki, and E. Oja. Independent component analysis in decomposition of auditory and somatosensory evoked fields. In *Proc. First International Conference on Independent Component Analysis and Blind Source Separation ICA'99*, pages 167–172, Aussois, France, January 11–15, 1999.

R. Vigário, V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja. Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In *Advances in Neural Information Processing Systems 10*. MIT Press, 1998.

R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja. Independent component approach to the analysis of EEG and MEG recordings. *IEEE Transactions on Biomedical Engineering*, 47(5):589–593, 2000.

G. Wübbeler, A. Ziehe, B.-M. Mackert, K.-R. Müller, L. Trahms, and G. Curio. Independent component analysis of non-invasively recorded cortical magnetic DC-fields in humans. *IEEE Transactions on Biomedical Engineering*, 47(5):594–599, 2000.

M. Zibulevsky and B. A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural Computation*, 13(4):863–882, April 2001.

A. Ziehe, K.-R. Müller, G. Nolte, B.-M. Mackert, and G. Curio. Artifact reduction in magnetoneurography based on time-delayed second order correlations. *IEEE Transactions on Biomedical Engineering*, 47(1):75–87, January 2000.

8 Blind Decomposition of Multimodal Evoked Responses and DC Fields

Ricardo Vigário[1]***, Andreas Ziehe, and Klaus-Robert Müller***
Fraunhofer Institut FIRST, Kekuléstr. 7, D-12489 Berlin, Germany
{rvigario, ziehe, klaus}@first.fhg.de

Gerd Wübbeler and Lutz Trahms
Physikalisch-Technische Bundesanstalt
Abbestr. 2-12, D-10587 Berlin, Germany
{Gerd.Wuebbeler, Lutz.Trahms}@ptb.de

Bruno-Marcel Mackert and Gabriel Curio
Neurophysics Group, Dept. of Neurology, Klinikum Benjamin Franklin,
FU Berlin, Hindenburgdamm 30, 12203 Berlin, Germany.
{macbm, curio}@zedat.fu-berlin.de

Veikko Jousmäki
Brain Research Unit, Helsinki University of Technology
P.O. Box 2200, FIN-02015 HUT, Finland
{veikko}@neuro.hut.fi

Jaakko Särelä and Erkki Oja
Neural Networks Research Centre
Helsinki University of Technology
P.O. Box 5400, FIN-02015 HUT, Finland
{Jaakko.Sarela, Erkki.Oja}@hut.fi

The advent of new brain mapping techniques, together with better and faster data storage capabilities, is generating a considerable amount of high-dimensional data. Suitable projecting or feature extraction mechanisms are required, able to reveal

1. To whom correspondence should be addressed. R.Vigário is on leave from Neural Networks Research Centre, HUT, P.O. Box 5400, FIN-02015 HUT, Finland; Ricardo.Vigario@hut.fi.

simple structures that may be easier to analyse than the complex brain activity that is often available to the physician, or brain researcher.

In data analysis we often face the following dilemma: if we impose a too strong model on the data, we might only get the structure that we are imposing; if our model is too weak we might get no useful result at all. As there is no systematic answer to this fundamental problem for all situations, we will discuss about possibilities and limits of the new blind source separation (BSS) technique in the context of specific biomedical applications. Here a fair amount of physiological and physics knowledge is available and we can use this prior information to bias our solution – of course carefully avoiding to predetermine the solution.

BSS methods, such as the ones based on independent component analysis (ICA) and temporal decorrelation (TD) methods have been shown to be an efficient tool for artifact identification and extraction from electroencephalographic (EEG) and magnetoencephalographic (MEG) recordings, as well as the analysis of some evoked and spontaneous brain activity.

This chapter reviews our recent results to the application of *blind* and *not so blind* source separation techniques to the analysis of evoked brain signals, elicited by sensory stimuli, and to the analysis of single trials of near DC brain fields.

8.1 Introduction

With the advent of new anatomical and functional imaging methods, it is now possible to collect, non-invasively, vast amounts of data from the living and active human brain. It has thus become very important to extract the essential features from the data to allow an easier representation or interpretation of their properties. Traditional approaches to solve this feature extraction or dimension reduction problem include, *e.g.*, principal component analysis (PCA), projection pursuit, and factor analysis. This chapter focuses on a signal processing technique, BSS, which allows the blind separation of sources, linearly mixed at the sensors.

It has turned out, with an assumption on statistical independence for the sources, that BSS very often provides the ideal "weak" model for decomposing brain signals like Electroencephalograms (EEG) or Magnetoencephalograms (MEG) since the assumption of independence is often verified (we will elaborate this aspect in more detail in Section 8.5). EEG and MEG are recordings of electric and magnetic fields of signals emerging from a multitude of neural currents within the brain. They are arguably the only existing completely non-invasive methods capable of giving direct information about the neural dynamics on a millisecond scale, which makes them attractive methods for functional brain imaging and diagnosis. The developments of EEG and MEG over the past years are strongly related to significant improvements in the quality of the sensing devices, the number of channels, the signal processing techniques and the neural source models used.

In order for neural magnetic fields to be measured outside of the head, the synchronous activation of tens of millions synapses is required. This limits the

spatial resolution of EEG and MEG to around 1 cm^2. Also, the synchronicity restrictions limits the number of macroscopically observable sources, active at a given time.

In Section 8.2 we present a short description of the blind source separation (BSS) problem. The independent component analysis (ICA) theory is then introduced, together with some algorithms capable of performing such analysis. In Section 8.4 another algorithm (TDSEP) is introduced which uses the temporal structure of the data to perform the blind separation of sources. In Section 8.5 we shortly validate the application of the ICA model to EEG and MEG data, and give an overview of the limitations of the BSS model. The use of ICA for the decomposition of event-related activity is presented in Section 8.7. In Section 8.8 the ICA approach is extended to incorporate existing prior modeling information in the search for the independent components. We conclude the chapter with the application of TDSEP to the identification of a DC-component from MEG recordings.

Due to the reviewing nature of this chapter, the expose herein is strongly based on previously published material (cf. (Vigário, 1999; Vigário et al., 2000; Vigário and Oja, 2000; Wübbeler et al., 2000; Ziehe et al., 2000a)).

8.2 The Blind Source Separation Model

Blind source or signal separation (BSS) is a very important problem in science and in engineering. It consists in revealing unknown sources from their linear mixtures, with very little, if any, knowledge on the mixing process. Only very few assumptions can be made on the source signals.

Let us assume that, at time instant k, the observed n-dimensional data vector, $\mathbf{x}(k) = [x_1(k), \ldots, x_n(k)]^T$ is given by the model:

$$x_i(k) = a_{i1}s_1(k) + a_{i2}s_2(k) + \cdots + a_{im}s_m(k)$$

or, in a more compact matrix notation,

$$\mathbf{x}(k) = \sum_{j=1}^{m} \mathbf{a}_j s_j(k) = \mathbf{A}\mathbf{s}(k). \tag{8.1}$$

The source signals, $s_1(k), \ldots, s_m(k)$, are unknown, as are the coefficients of the mixing matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_m]$. The goal is therefore to estimate both unknowns from a sample of the $\mathbf{x}(k)$. The solution is sought in the form

$$\mathbf{y}(k) = \hat{\mathbf{s}}(k) = \mathbf{B}\mathbf{x}(k), \tag{8.2}$$

where $\mathbf{B}$ is called the separating matrix.

The general BSS problem requires $\mathbf{A}$ to be an $n \times m$ matrix of full rank, with $n \geq m$ (*i.e.*, there are at least as many mixtures as the number of independent sources). In most algorithmic derivations, an equal number of sources and sensors is assumed.

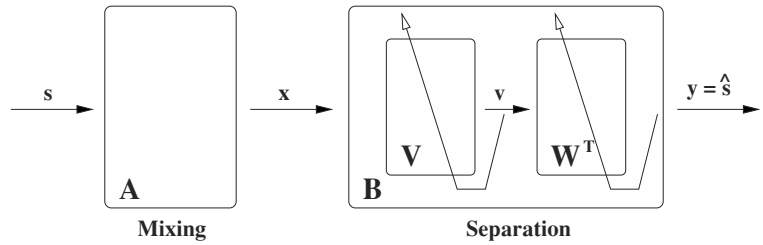


Figure 8.1 Schematical illustration of the mathematical model used to perform the ICA decomposition.

In the model summarised by Eq. (8.1), and schematically illustrated in Fig. 8.1, we omit any additive noise. Some considerations on noisy models can be found in (Hyvärinen, 1999; Vigário, 1999; Müller et al., 1999; Kawanabe and Murata, 2000; Särelä et al., 2001).

It is important to note that, in our application of BSS approaches to the analysis of EEG or MEG recordings, the estimates of both the underlying source signals and the mixing matrix are of importance. The former will give us information on the temporal activation of a particular brain source, whereas the latter will give the corresponding field patterns, which are required for the localization of the brain activation areas.

8.3 Independent Component Analysis

ICA is a novel statistical technique that aims at finding linear projections of the data that maximize their mutual independence. Its main applications are in feature extraction and blind source separation.

ICA attempts to solve the BSS problem by assuming that the underlying source signals $\mathbf{s}$ are stationary, independent and with only up to one source allowed to have a Gaussian probability distribution.

It is not the intention of the authors of this chapter to fully characterise the field of ICA. A good and recent review of the theory behind ICA and several of its applications can be found in (Hyvärinen et al., 2001). Yet, in order for the reader to understand the following application oriented sections, we believe it necessary to spend some time in the description of the algorithms used therein.

8.3.1 The FastICA Algorithm

In the FastICA algorithm (Hyvärinen and Oja, 1997), to be described below, the initial step is whitening or sphering (see Fig. 8.1 for an illustration of the notation used). By a linear transformation, the measurements $x_i(k)$ and $x_j(k)$, for all i, j, are made uncorrelated and unit-variance. The whitening facilitates the separation

of the underlying independent signals (Hyvärinen et al., 2001). Furthermore, Särelä and Vigário (2001) have shown that a well chosen compression, during this stage, may be necessary in order to reduce the overlearning(overfitting), typical of ICA methods.

The whitening may be accomplished by PCA projection: $\mathbf{v}(k) = \mathbf{V}\mathbf{x}(k)$, with $E\{\mathbf{v}(k)\mathbf{v}(k)^T\} = I$. The whitening matrix $\mathbf{V}$ is given by $\mathbf{V} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^T$, where $\mathbf{\Lambda} = \text{diag}[\lambda(1), \ldots, \lambda(m)]$ is a diagonal matrix with the eigenvalues of the data covariance matrix $E\{\mathbf{x}(k)\mathbf{x}(k)^T\}$, and $\mathbf{U}$ is a matrix with the corresponding eigenvectors as its columns. The transformed vectors $\mathbf{v}(k)$ are called white or sphered, because all directions have the same unit variance.

In terms of $\mathbf{v}(k)$, the model (8.1) becomes $\mathbf{v}(k) = \mathbf{VA}\mathbf{s}(k)$, and we can show that the matrix $\mathbf{W} = \mathbf{VA}$ is orthogonal (Hyvärinen et al., 2001). Therefore, the solution is now sought in the form:

$$\hat{\mathbf{s}}(k) = \mathbf{W}^T\mathbf{v}(k). \tag{8.3}$$

The decorrelation attained with a PCA decomposition is not enough when dealing with the more restrictive independence requirements. Methods based on higher order statistics are required. Their theoretical foundations can be tracked to notions such as maximization of negentropy, non-Gaussianity or minimization of mutual information.

In many ICA algorithms, the fourth-order cumulant also called the kurtosis is used as a measure of non-Gaussianity. For the ith source signal, the kurtosis is defined as $kurt(s_i) = E\{s_i^4\} - 3[E\{s_i^2\}]^2$. $E\{\cdot\}$ denotes the mathematical expectation value of the bracketed quantity. The kurtosis is negative for source signals whose amplitude has sub-Gaussian probability densities (distributions flatter than Gaussian), positive for super-Gaussian (sharper than Gaussian, and with longer tails), and zero for Gaussian densities. Maximizing the norm of the kurtosis leads to the identification of non-Gaussian sources.

Consider a linear combination $y = \mathbf{w}^T\mathbf{v}$ of the white random vector $\mathbf{v}$, with $\|\mathbf{w}\| = 1$. Then $E\{y^2\} = 1$ and $kurt(y) = E\{y^4\} - 3$, whose gradient with respect to $\mathbf{w}$ is $4E\{\mathbf{v}(\mathbf{w}^T\mathbf{v})^3\}$.

The FastICA (Hyvärinen and Oja, 1997) is a fixed point algorithm which finds one of the columns of the separating matrix $\mathbf{W}$ (noted $\mathbf{w}$) and so identifies one independent source at a time. The corresponding independent source signal can then be found using Eq. (8.3). Each lth iteration of this algorithm is defined as

$$\begin{aligned} \mathbf{w}^*{}_l &= E\{\mathbf{v}(\mathbf{w}_{l-1}^T\mathbf{v})^3\} - 3\mathbf{w}_{l-1} \\ \mathbf{w}_l &= \mathbf{w}^*{}_l/\|\mathbf{w}^*{}_l\|. \end{aligned} \tag{8.4}$$

In order to estimate more than one solution, and up to a maximum of m, the algorithm may be run repeatedly. It is, nevertheless, necessary to remove the information contained in the solutions already found, to estimate a different independent component each time. For details on the FastICA algorithm, see (Hyvärinen and Oja, 1997). Further reading on algorithmic implementations of the ICA technique,

as well as its extensions and relations to other data analysis techniques, can be found, e.g., in (Vigário, 1999) and (Hyvärinen et al., 2001). A Matlab^{TM} package that performs the FastICA can be found at:
`http://www.cis.hut.fi/projects/ica/fastica/`.

8.3.2 INFOMAX

Another approximation to the ICA decomposition can as well be attained by maximizing the output entropy, or information flow, of a neural networks with nonlinear outputs. Bell and Sejnowski (1995) proposed one such algorithm that maximizes the mutual information I between the inputs and the outputs of a neural network. The intuitive interpretation of $I(\mathbf{x}, \mathbf{y})$ is the reduction of information in $\mathbf{x}$, after the observation of $\mathbf{y}$.

The separating matrix $\mathbf{B}$ is found using the updating rule:

$$\Delta\mathbf{B} \propto (\mathbf{I} - 2\tanh(\mathbf{y})\mathbf{y}^{\mathbf{T}})\mathbf{B}. \tag{8.5}$$

Note that this algorithm does not require an explicit pre-whitening. A Matlab^{TM} package that performs the INFOMAX can be found at:
`http://www.cnl.salk.edu/~scott/ica-download-form.html`.

8.3.3 JADE

For off-line (batch) computation, Cardoso and Souloumiac (1993) developed the JADE algorithm based on the (joint) diagonalization of matrices obtained from 'parallel slices' of the fourth-order cumulant tensor. This algorithm often performs very efficiently on low dimensional data if sufficiently many sample points are available. However, for high dimensional problems like MEG the effort for storing and processing the 4-th order cumulants is $\mathcal{O}(m^4)$ and computation may become prohibitive. A Matlab^{TM} package that performs the JADE can be found at:
`ftp://sig.enst.fr/pub/jfc/Algo/Jade`.

Several papers have been written on the relations between the active principles behind the FastICA, the Infomax, and the JADE algorithms. For a well structured comparison see, e.g., (Hyvärinen et al., 2001).

8.4 Temporal Decorrelation Methods

In addition to the ICA approach to source separation, described in the previous section, there exist other useful criteria to define a suitable decomposition. In particular, EEG and MEG recordings have a rich dynamical time structure. Therefore, instead of basing our decomposition approach on the distributional information contained in the data, one can exploit directly temporal, i.e., spectral information, to perform the desired blind separation of sources.

An attractive framework to deal with this problem is the simultaneous diagonalization of appropriately defined matrices by algebraic methods (cf. (Tong et al., 1991; Molgedey and Schuster, 1994; Belouchrani et al., 1997; Ziehe et al., 1998; Wu and Principe, 1999; Ziehe et al., 2000b)). In particular, the TDSEP algorithm by Ziehe and Müller (1998) finds an estimate of the mixing matrix $\mathbf{A}$ by simultaneously diagonalizing several time lagged correlation matrices $R_{\tau(\mathbf{x})} = \langle \mathbf{x}(t)\mathbf{x}^T(t-\tau)\rangle$.

This method can be seen as an efficient way to minimize the cost function

$$J(W_{ij}) = \sum_{l=1}^{N} \sum_{i \neq j} \langle y_i(k) y_j(k+\tau_l)\rangle^2, \tag{8.6}$$

that measures the correlation of the outputs for several time-lags τ_l.

As with the ICA algorithm presented in Section 8.3.1, the simultaneous diagonalization procedure that is employed here consists of a sphering or whitening step followed by a rotation. The determination of the rotation matrix $\mathbf{W}$, is performed by a Jacobi-type method as in (Cardoso and Comon, 1996), applied to the set of time-delayed correlation matrices.

From the computational point of view TDSEP is very efficient and robust, since (1) it uses linear algebra and avoids complicated optimizations and (2) it relies on estimates of simple time-lagged covariance matrices (second-order statistics). Therefore TDSEP is ideally suited for the preprocessing of multi-channel data typically encountered in physiological recordings. A Matlab^{TM} implementation of this algorithm can be found at:
`http://www.first.fhg.de/~ziehe/research.html`.

8.5 On the Validity of the Linear ICA Model

The application of ICA to the study of EEG and MEG signals assumes that several conditions are verified: the existence of statistically independent source signals, their instantaneous linear mixing at the sensors, and the stationarity of both the source signals and the mixing process.

8.5.1 Instantaneous Linear Mixing

Because most of the energy in EEG and MEG signals lies below 1 kHz, the quasistatic approximation of the Maxwell equations holds, and each time instance can be considered separately (Hämäläinen et al., 1993). Therefore, the instantaneous mixing model is valid. The linearity of the mixing follows as well directly from the Maxwell's equations.

8.5.2 The Independence

The independence criterion applies solely to the statistical relations between the amplitude distributions of the signals involved, and not to considerations upon the morphology or physiology of certain neural structures. In particular, the obvious time relations between stimuli and brain responses (which correspond to a certain form of dependence), have no influence on the statistical independence between two signals with factorizable joint probability density.

On the other hand, in certain conditions, the search for independent components can be replaced by a search for maximally non-Gaussian linear transformation of the data. Due to their sparse activation patterns, the evoked responses are clear examples of non-Gaussian distributed signals. This means that, even though the direct independence criterion may occasionally be difficult to justify, the ICA model may still be useful.

Rhythmic activity in the brain poses an additional problem. Pure oscillatory activity has negative kurtosis. Yet in reality neural activity often comes as bursts of limited time span. Depending on the lengths of these bursts, the sign of the kurtosis may be positive. In the worst case, the global kurtosis may even be zero, *i.e.*, the desired component is then interpreted as Gaussian by any kurtosis-based method. A different strategy may be required to cope with this problem.

The TDSEP algorithm, described in Section 8.4, uses explicitly the temporal correlations present in the data, and is therefore very suitable to handle brain oscillation data.

Furthermore, Barros et al. (2000) show that a modified version of the FastICA, giving particular focus to periodic or quasiperiodic signals, deals well with cardiac and other periodic contaminations of MEG recordings. The same approach may lead to very interesting results when analysing the brain's rhythmic activity.

One can also model the dynamics of the oscillatory sources directly. In (Särelä et al., 2001), the sources are modelled by MLP networks, resulting in the dynamical factor analysis (DFA). There, the overfitting, typical in maximum likelihood or maximum a posteriori estimates of such a complex model, have been avoided using a Bayesian based approach called ensemble learning (Hinton and van Camp, 1993; Lappalainen and Miskin, 2000). As an additional asset, the Bayesian approach makes it simple to estimate the noise in the observations.

8.5.3 Stationarity

The last requirement for the utilization of the basic ICA model is the stationarity of the recordings. This requirement applies both to the source signals, and to the mixing model. Stationarity of the independent source signals is required to assume the convergence of their amplitudes to a particular distribution. The stationarity in the mixing model ensures a constant mixing matrix $\mathbf{A}$.

Generally, the non-stationarity of EEG and MEG signals is well documented (Blanco et al., 1995). Yet, in the implementation of the batch FastICA algorithm,

the estimation of the distribution of the independent signals is made using the whole data set, removing the need for stationarity requirements.

Another way to tackle the intrinsic non-stationarity of EEG and MEG signals has been suggested in (Müller et al., 2000). There, an "annealing competition of experts" is used to perform the segmentation of non-stationary EEG signals into stationary periods. Any BSS approach can then be applied to each period.

The stationarity of the mixing model corresponds to the assumption of fixed field patterns associated with the different independent components. In the widely used dipole source model, the mixing stationarity leads to the existence of sources with fixed locations and orientations, but amplitude varying with time. Such models (Scherg and von Cramon, 1985; Mosher et al., 1992) have been extensively and efficiently used in the analysis of evoked responses, which justifies the use of constant mixing vectors $\mathbf{a}_i$ in our BSS model.

If strictly required, also ICA algorithms with non-stationary mixing models exist (see, e.g., (Murata et al., 1997)).

8.6 Limits and Problems of Source Separation Algorithms

It is important to note that the blind separation of sources is ultimately bound to the acting principle employed. This means that the orthogonality of the decomposition is the most we can get from PCA, the orthogonality in some feature space what we get from kernel PCA (Schölkopf et al., 1998) and the mutual independence of the components when using ICA.

The efficiency of high-order statistics to produce good ICA algorithms (the kurtosis based FastICA, JADE,...) comes with a clear price: the strong sensitivity to outliers. In fact, these often turn out to be the leading factors in the search for the underlying sources.

Another limitation of the presented BSS approaches is that one can extract only up to as many underlying sources as the number of sensors. In the brain, however, the multitude of microscopic sources outnumbers by far the number of sensors. Nevertheless, as stated in the introductory section, it can be argued that the total number of macroscopically observable sources, active at a given time, is of a much smaller count.

On a more general note, the linear mixing model, summarized in Eq. (8.1), may be too simple to accommodate possible nonlinearities in the data or even convolutive mixtures. If we have shown that the instantaneous linear mixing is a good approximation for EEG and MEG data, that is not the case in the temporal dynamics of the neural activations. A non-linear model of such processes may be necessary if a proper temporal modeling is sought. DFA is one example of algorithms that perform such non-linear modeling (Särelä et al., 2001).

8.6.1 The Necessary Amount of Data: Overlearning and Overfitting

When the number of samples of the observed data is insufficient to explain the high-dimensional independent source signals, we may encounter overlearning or overfitting effects in ICA algorithms. This result is true for practically any type of contrast-based linear ICA implementations (Särelä and Vigário, 2001). In the extreme case of an equal number of independent source signals, mixtures and sample sizes, it is shown that the optimal solution is the one that produces source signals that are zero almost everywhere except for a single spike or bump. In the referred communication, experimental evidence of this overfitting effect is given, using both simulated and real MEG data.

Temporal decorrelation algorithms, such as the TDSEP, may suffer from a different type of overlearning, consisting in a series of sinusoidal components of various frequency contents.

Furthermore, channel noise, explicitly ignored in the model of Eq. (8.1), effectively doubles the number of independent sources, rendering the overlearning problem ever more present. In some applications, we may construct an approximate noise model. Projections to signal spaces orthogonal to the noise space could then be performed (Hyvärinen, 1998; Müller et al., 1999; Hyvärinen et al., 2001; Särelä et al., 2001).

When utilizing algorithms requiring pre-whitening, this preparatory stage can be used to reduce the dimensionality of the data. The compression, specially if the rejected components contain mainly noise (i.e. unnecessary information), should reduce the overlearning effect due to lack of net information.

Ultimately, a good reliability measure of the components found may help to determine whether we are in presence of clear overlearning, or of an overlearning-like true source. Furthermore, such measure could help determining whether the model is appropriate at all. One such measure, based on a bootstrap resampling approach has been recently proposed by Meinecke et al. (2002).

8.7 Analysis of Multimodal Evoked Fields

State-of-the-art approaches for processing magnetic evoked fields are often based on a careful expert scrutiny of the complete data (in raw format or averaged around the responses to repeating stimuli). At each time instance, one or a set of neural sources, often of dipolar nature, are modeled in order to produce as good fit to the data as possible. The quality of the fit is then evaluated through its goodness-of-fit (Kaukoranta et al., 1986), a normalized squared error between the measurements and the fields produced by the modeled sources:

$$g = \left(1 - \frac{\sum_i (\mathbf{b}_i - \mathbf{m}_i)^2}{\sum_i \mathbf{m}_i^2}\right) \times 100\%.$$

The summations run over the complete set of sensors. The higher the g, the better the explanation. If $g = 100\%$, the model fully explains the measurements, if $g = 0\%$, it does not perform any better than a zero field fit would.

The choice of the time instances where the fitting of the model should be made, as well as the type of source models employed, are crucial. An approach based in ICA presents the advantage of constant field patterns, simplifying the analysis of the results, and enabling a semi-automatic processing of the data.

The application of ICA in event related studies was first introduced in the blind separation of auditory evoked potentials by Makeig et al. (1997). This method has been further developed in relation to magnetic auditory and somatosensory evoked fields (AEFs and SEFs, respectively) in (Vigário et al., 1998, 1999), using the FastICA algorithm, in a deflationary mode.

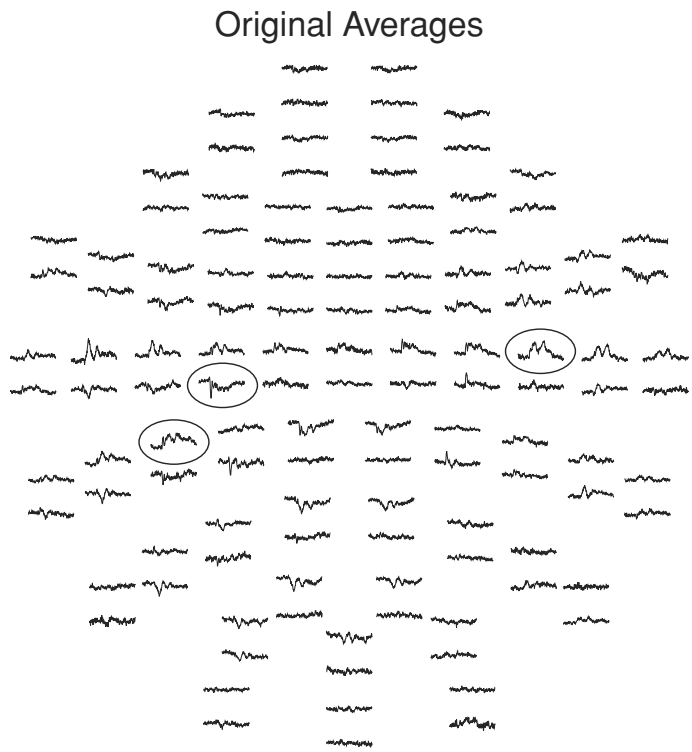


Figure 8.2 An example of evoked MEG signals to concomitant vibrotactile and auditory auditory stimulation. The sensor array of the neuromagnetometer (Neuromag-122 (Hämäläinen et al., 1993)) is viewed from the top, i.e. left hemisphere is on the left side and right hemisphere on the right side, and the nose of the subject is pointing upwards. Each trace corresponds to the signal which was detected by a single sensor as a function of time. Encircled channels show the maximum responses to auditory and tactile stimulation. From (Vigário and Oja, 2000).

Without any prior source model assumption, other than the statistical independence to the rest of the MEG signals, the most significant independent components we have found in different modality event-related studies have shown patterns that agree with the conventional dipole approximation. Adding the dipole modeling information to the calculation of the source locations, we have found them to fall on very acceptable brain areas (the difference to conventional methods was well below 1 cm, therefore within the spatial precision of MEG).

In (Vigário et al., 1999), ICA was shown to be able to differentiate between somatosensory and auditory brain responses in the case of vibrotactile stimulation. The stimuli was generated with a bass-reflex loudspeaker and a tube delivering the tone to a balloon which was held by the subject with both hands. The sound pressure level of the simulator was about 60 dB SPL, and thus somatosensory evoked fields (SEFs) to vibrotactile stimuli (200 Hz, duration 100 ms) and auditory evoked fields (AEFs) to the concomitant auditory stimulation were elicited in the same experiment (Jousmäki and Hari, 1999).

Figure 8.2 shows the complete set of 122 averaged evoked responses. Three inserts in the figure highlight equal number of interesting signal types present in the measurements. On the leftmost one, we can see a step-like signal. The middle one shows a sharp and early response, originated in the primary somatosensory cortex. The rightmost insert has a broader signal, with longer latency than the previous one, which is associated with the primary auditory responses.

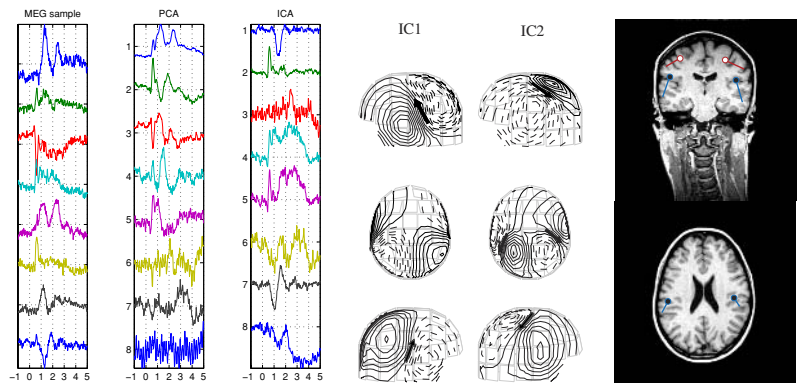


Figure 8.3 Results of the application of FastICA to averaged brain MEG responses to a vibrotactile stimulation. Frames a) through c) present, respectively, a sample of the original MEG data, the whitened and the independent signals. Each tick corresponds to a time interval of 100 ms. Frame e) shows the field patterns associated with the first two independent components and the modeled current dipoles (arrows). Their corresponding location in the brain is showed in frame e) superimposed to magnetic resonance images. From (Vigário and Oja, 2000).

In Fig. 8.3, together with a sample of the MEG averages, are depicted the first 8 principal components, as well as the first 8 components found on a single run of the FastICA algorithm. Note that PCA hasn't been able to resolve the complexity of the original MEG signals, most of the components presenting still combined somatosensory and auditory responses (see Fig. 8.3b). The independent components in 8.3c show a much improved separation.

The field patterns shown in Fig. 8.3d, correspond to the columns of the estimated mixing matrix associated with the first two independent components. Full lines depict the magnetic flux coming out of the head, whereas dashed ones correspond to the flux entering the head. The current dipoles best accounting for these field patterns are also shown.

The location of the equivalent current sources fall on the expected brain regions for the particular stimulus. Fig. 8.3e shows these brain sources superimposed onto vertical and horizontal MRI slices. The black dots in the MRI correspond to activation of the primary auditory cortex. In addition, the orientations of the dipoles, represented by the lines starting at the dots, are approximately normal to the surface of the cortex. The white dots, based on the second independent component, correspond to the activation on the primary somatosensory area.

8.8 Extending the Basic ICA and BSS

8.8.1 How B Should BSS Be?

Thus far we have always used ICA in an attempt to extract, with minimal *a priori* information, independent source signals from their instantaneous linear mixtures. Yet, we often know more than what we are using. In fact, when attempting to give a physiologically plausible explanation to the independent components found from the evoked responses, we have admitted a dipolar current model for the corresponding neural sources.

Knuth (1998) proposes a Bayesian framework, in which it is possible to incorporate, onto the ICA/BSS model, prior information about the source geometry and the mixing properties.

In the next section we show some preliminary results on the incorporation of a dipolar model flavor to the ICA framework (Vigário, 2000). We thus reduce the blindness of the search, and at times relax the independence assumption.

8.8.2 The Dipole Modeled FastICA

The study presented in this section consists of a modification on the FastICA algorithm, in order to add an explicit stage of dipolar modeling into the search for the independent components. This iterative method is applied to the decomposition of evoked fields introduced in Section 8.7.

8.8.2.1 From the Original FastICA

FastICA can be used both in a symmetrical and a deflationary manner. If the independence is not fully guaranteed, the error of assuming it will be somewhat spread to all independent components in the symmetrical version. In the deflationary, the early components are more likely to fit the criterion, as the later ones will have to deal with accumulated errors from the previous components found.

In the present work, FastICA will be used in its deflationary version, to extract one single component. The independence of this component should then be the most reliable one. Yet, the order of appearance of the independent components is undetermined in ICA, therefore different initial conditions can lead to different first guesses from the algorithm. Due to the very fast convergence of the FastICA, as much as 100 different initial conditions were tested, and their results gathered. Only a few different signals were consistently picked using this method, as can be seen in Fig. 8.4.

8.8.2.2 Adding Dipole Modeling to FastICA

As stated above, dipole modeling is often assumed when validating the independent components and locating the corresponding neural sources. In fact, often the field patterns, mapping the components to the measurements, fit well the dipole modeling assumptions (see, e.g., Fig. 8.3d).

Furthermore, in the deflationary FastICA, to extract more than one independent component, we have to insure that the contributions of the ones already found are extracted from the original data. This can be done by explicit subtraction of the independent component, or by imposing the orthogonality of the spaces defined by successive components.

We now propose to remove, not the independent component itself, but the magnetic field produced by one or more equivalent current dipoles (ECDs) that fit the best the respective independent component. With this change, the independence criterion is somewhat relaxed, but with an increase in the explanation power of the resulting component. This modified or iterated FastICA algorithm is therefore a good compromise between pure independence and conventional ECD fitting. Or, in other words, between the weak BSS model, given by ICA, and the strong dipole-based modeling.

8.8.2.3 Results

Figure 8.4 shows the results of the ICA decomposition of the evoked responses. The left column has the four independent components found from 100 runs of the original FastICA on the complete data set, changing the initial conditions after each run. These results agree with the ones presented earlier, and explain well some of the measured signals of Fig. 8.2: OrIC1 and OrIC3 (first and third components extracted using the original FastICA algorithm) have latencies that

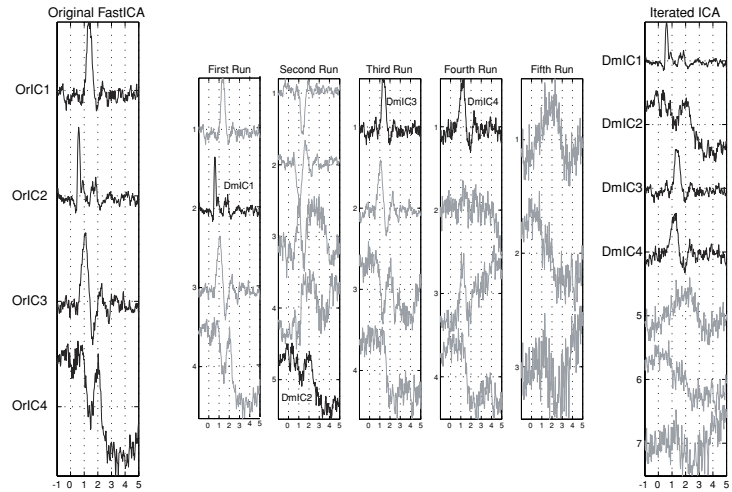


Figure 8.4 ICA decomposition of the evoked responses. On the left are depicted the independent components corresponding to a single run of FastICA, whereas on the right are the ones corresponding to the modified algorithm. The middle 5 columns show the construction of the components of the right hand side. From (Vigário and Oja, 2000).

are characteristic of the auditory responses visible in the rightmost insert, whereas OrIC2 and OrIC4 are clearly related to the somatosensory and step-like inserts.

Note as well that the multiple extraction of only one independent component leads to the identification of only the most interesting components found in Fig. 8.3. This consistency increases our trust in this ICA decomposition.

The locations of the ECDs associated to the independent components are shown in Fig. 8.5. Once again, it is clear that they agree with the expected brain regions to be activated by the particular stimulus mode. Note that only the last component required three dipoles to correctly explain the respective independent component. One of the ECDs explaining OrIC4 originates from the primary somatosensory area, already used to explain OrIC2.

The independent components shown in the right column of Fig. 8.4 have been produced by the modified FastICA algorithm. After each set of 100 runs of FastICA, the most frequently observed independent component was kept. The magnetic field associated to its best ECD fit was then extracted from the original data. The nth independent component is now sought from a set of 122 dimensional signals, from which the ECD contributions of the previous $n-1$ components have been removed. The 5 columns in the middle show each step of the long iterative process leading to the final set of signals on the right. Note that, on the second run, the signal picked was not the one corresponding to the auditory response, as we could have wished for, but rather the step-like one. This is due to the fact that the choosing criterion has no way to evaluate the goodness or physiological usefulness of each component.

Figure 8.6 shows the field patterns of the independent components found using the dipole modeled FastICA algorithm, and respective ECDs. If it is clear that

DmIC1 and DmIC3 (first and third components using the dipole modeled version of FastICA) present great resemblance to OrIC2 and OrICA1, the same can not be said from the other independent component. Due to an early extraction of the somatosensory ECDs, no such dipole was required to explain DmIC2, reducing here the total number of dipoles from 9 to 8. Furthermore, the pattern on the left side of DmIC2 is now much clearer than that of OrIC4. Finally, it is as well visible that DmIC4's patterns are simpler than those of the corresponding original decomposition (OrIC3).

In Fig. 8.7 we see the goodness-of-fit values attained by the complete sets of ECDs associated to both ICA implementations. The solid line shows the combined performance of the 9 dipoles modeling the "conventional" FastICA components, whereas the dashed line shows the performance of the 8 dipoles associated to the iterated FastICA decomposition. It is clear that the modified ICA algorithm explains the measured data at least as well as the traditional algorithm does, with increased performance on occasions (see, e.g., the portion corresponding to the activation of the primary auditory area).

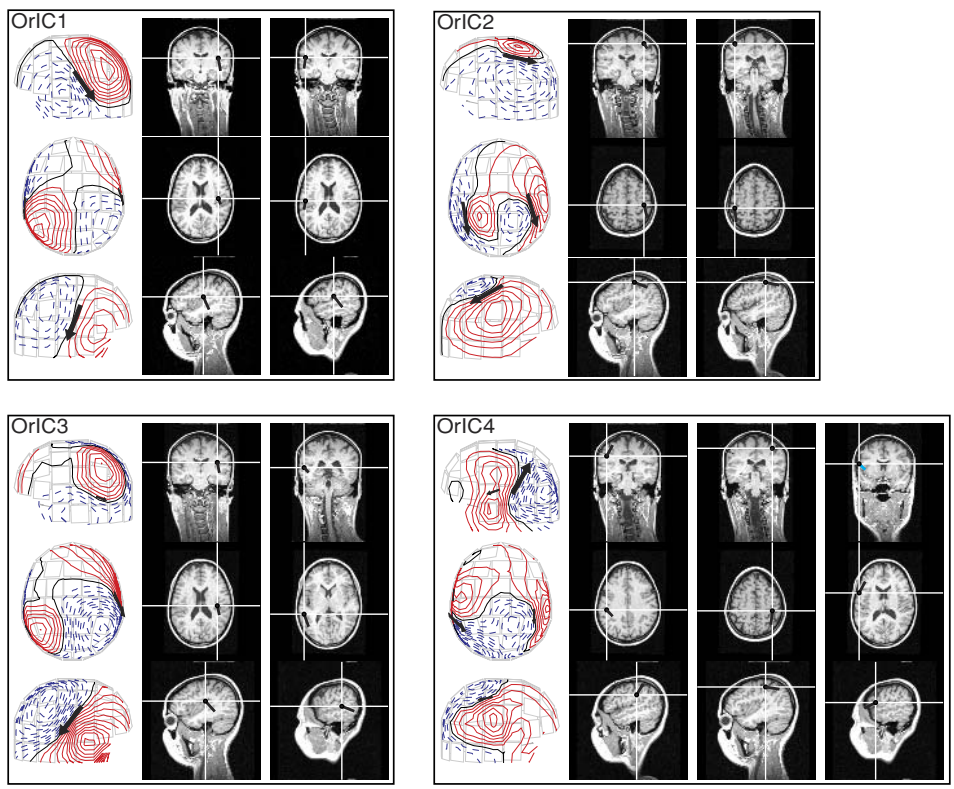


Figure 8.5 Localizations of the ECDs corresponding to the original FastICA decomposition. From (Vigário and Oja, 2000).

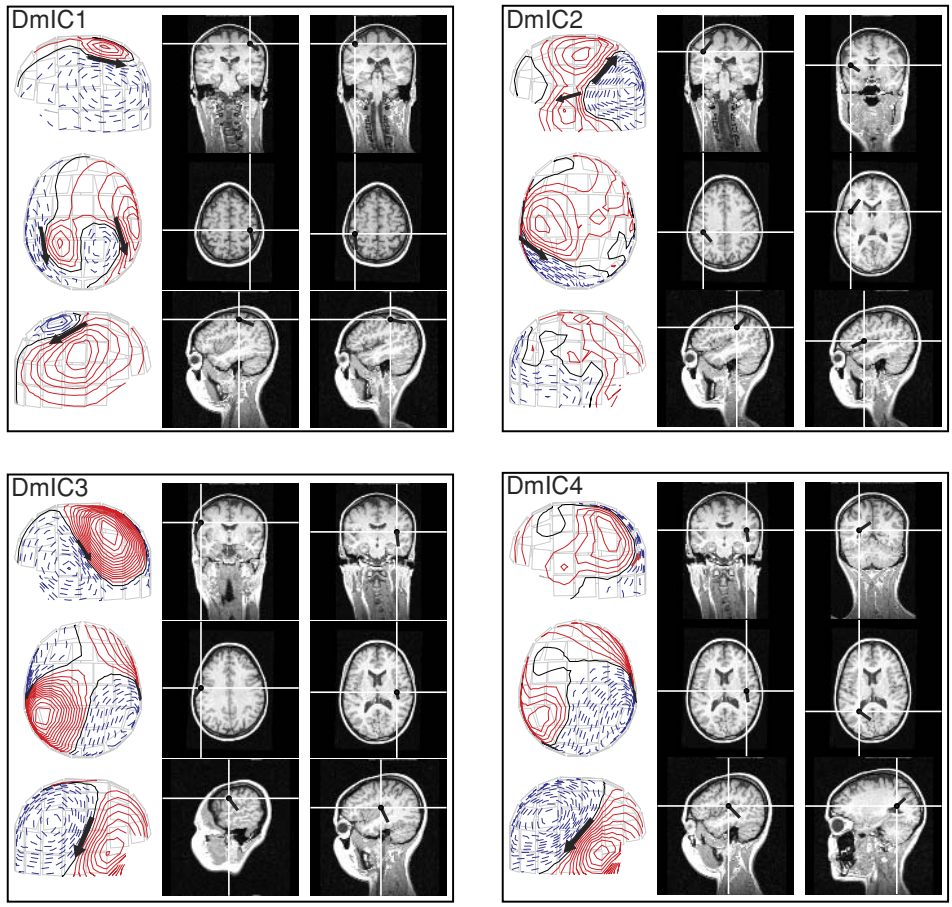


Figure 8.6 Localization of the modified FastICA. From (Vigário and Oja, 2000).

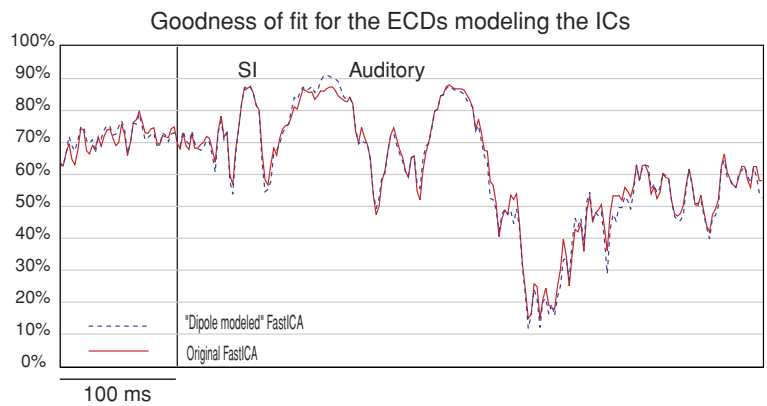


Figure 8.7 Goodness-of-fit for the ECDs associated with the two FastICA algorithms. From (Vigário and Oja, 2000).

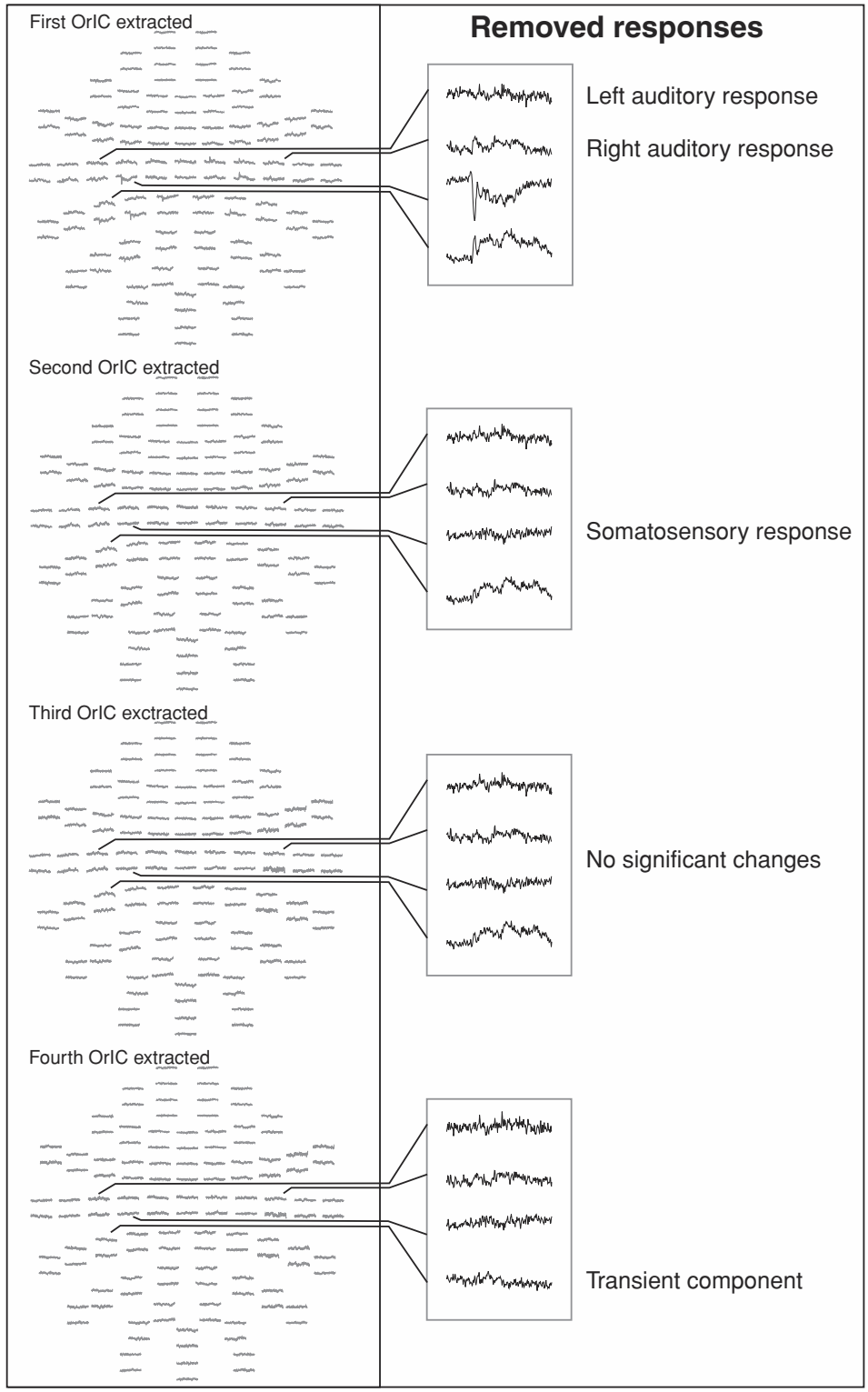


Figure 8.8 Residual magnetic fields, after extracting each "conventional" independent component's contribution. Adapted from (Vigário and Oja, 2000).

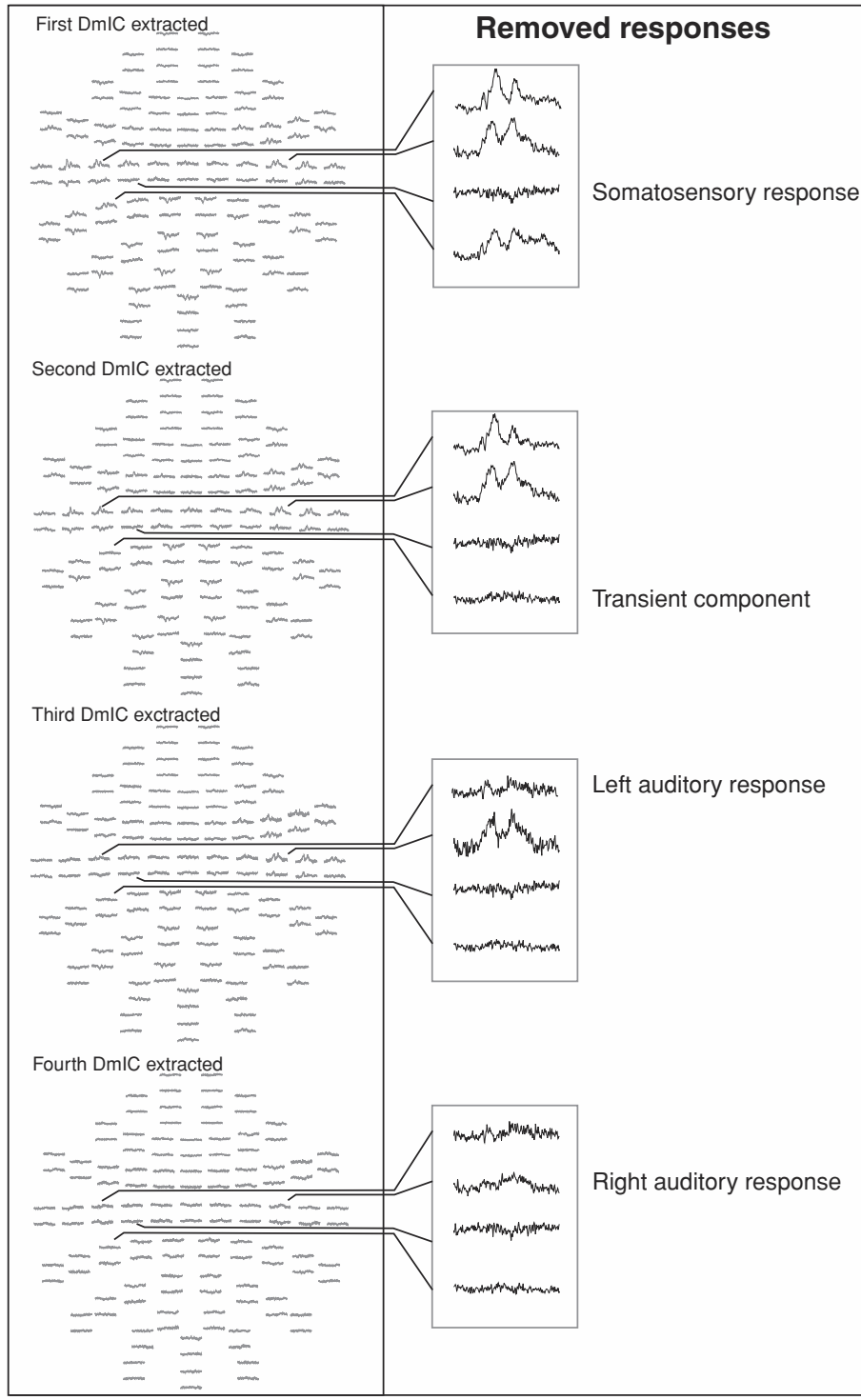


Figure 8.9 Residual magnetic fields, after extracting each iterated independent component's contribution. Adapted from (Vigário and Oja, 2000).

Another way to look into the different behaviors of the two algorithms, is through the analysis of the remaining measured signals, after extracting the magnetic fields originated by the ECDs corresponding to each independent component.

In Fig. 8.8, we show these remaining magnetic fields for the conventional FastICA. Symmilarly, Fig. 8.9 shows the remaining fields for the modified algorithm. On the original FastICA results, the extraction of the first independent component leads to the suppression of all auditory related components (left and right). Removing the second component leads to a very simplified set of remaining signals, that seems to change only after the extraction of the fourth component, in which the step-like signal disappears.

In the modified algorithm (Fig. 8.9), after extracting the first independent component from the original data, the somatosensory related signal is suppressed. The step-like response is removed after the second. It is important to note that the extraction of the third independent component doesn't result in a complete suppression of the auditory responses. In fact, the right-hand side responses are just attenuated at this stage, and removed only after the extraction of the fourth independent component. This is a good indication that the algorithm is actually capable to detect the lateralization effects visible when the auditory stimuli are applied mainly to one ear as seen earlier (Vigário et al., 1998).

8.9 Cortical Magnetic DC Fields in Humans

Recently, advanced biomagnetic recording technology has opened the possibility to reliably detect slow electrophysiological processes, occurring over several seconds (Wübbeler et al., 1998).

Such near-DC phenomena are expected in metabolic injuries to brain cells in stroke or migraine, e.g. in anoxic depolarization, peri-infarct depolarization or spreading depression (Gardner-Medwin et al., 1991; Chen et al., 1992; Back et al., 1994). Non-invasive electrical recordings of near-DC phenomena are prone to large drift artifacts due to electrochemical instabilities at the electrode-skin interface. Up to now this limitation could be overcome only by invasive approaches (Hotary et al., 1992; Stys et al., 1991). In contrast, Superconducting Quantum Interference Devices (SQUIDs) in combination with a specialized mechanical modulation technique allow for a non-invasive registration of near-DC (below 0.1 Hz) magnetic fields. Biomagnetic fields in this frequency range were detected, quantified and continuously monitored in the human brain. This was achieved by employing an acoustical stimulation paradigm (alternating periods of music and silence, each of 30 s length, to the subjects right ear during 30 min. of total recording time) to induce a prolonged auditory cortex activation (for a detailed physiological background see (Mackert et al., 1999)).

8.9.1 Data Acquisition and Validation

In order to observe very low frequency brain activity, it is necessary to suppress the influence of DC and near-DC magnetic field noise. This can be achieved by careful hardware design and/or signal processing.

Therefore, the DC magnetic field values were acquired by using a mechanical horizontal modulation of the body position with a frequency of 0.4 Hz and an amplitude of 75 mm. This modulation transposed the DC magnetic field of the subject to the modulation frequency, which is less contaminated by noise. The recorded magnetic field data were processed by digital lock-in techniques in order to extract the modulation induced frequency components (Wübbeler et al., 1998). Then the DC-field of the subject was reconstructed from these frequency components by using a transformation technique based on a virtual magnetic field generator (Mackert et al., 1999). These reconstructed DC magnetic field values, sampled at the modulation frequency of 0.4 Hz, gave a total number of 720 sample points per channel for the 30 minutes recording time.

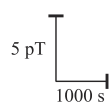


Figure 8.10 Input data used for BSS after DC preprocessing (demodulation and reconstruction); arranged according to sensor positions; diameter of sensor array 210 mm. From (Wübbeler et al., 2000).

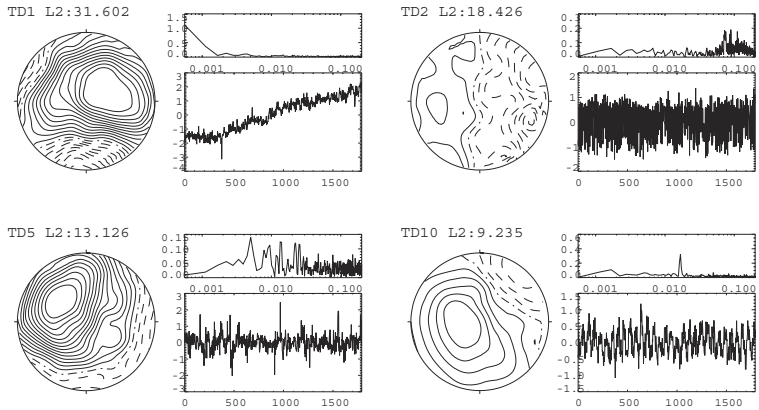


Figure 8.11 Spatial field patterns, waveforms and frequency contents of four selected components obtained by TDSEP. For units and details of ICA10 cf. Fig. 8.12. From (Wübbeler et al., 2000).

Examining this data (see Fig. 8.10), we observe that the signals have an obvious trend behavior (slow drift). Possible components of interest are covered by other strong signals of unknown origin, i.e. the very weak response to the stimulus is completely hidden.

8.9.2 Results and Discussion

The data was reduced to a 32-dimensional subspace, during whitening, prior to the application of TDSEP. In the latter algorithm, 50 time-lagged correlation matrices ($\tau = 1...50$ sample points) were used for simultaneous diagonalization. In Fig. 8.11 some selected components are shown. Not surprisingly, the first component (TD1) mainly captured the slow drift, already visible in the data (see Fig. 8.10). While most other components show irregular time courses, reflecting the dynamics of undetermined processes, it is noteworthy that their field maps feature spatially coherent field patterns which clearly distinguish them from random channel noise patterns.

Remarkably, one component (TD10) shows a (noisy) rectangular waveform. Its time course and frequency (see Fig. 8.12) clearly displays the $\frac{1}{30s}$ "on/off" characteristics of the stimulus. The spatial field distribution of TD10 shows a bipolar pattern. This is very similar to the field pattern obtained by classical averaging and subtraction of the averaged "off" periods as baseline (Mackert et al., 1999).

The field pattern of TD10 agrees with the N100m obtained, for the same subject, through conventional non-DC MEG recordings. Furthermore, its brain source is located at the expected position for the corresponding cortical activity (Hari et al., 1980; Mackert et al., 1999). Both findings regarding the time course and the field pattern give direct evidence that TD10 represents the response to the acoustical stimulus.

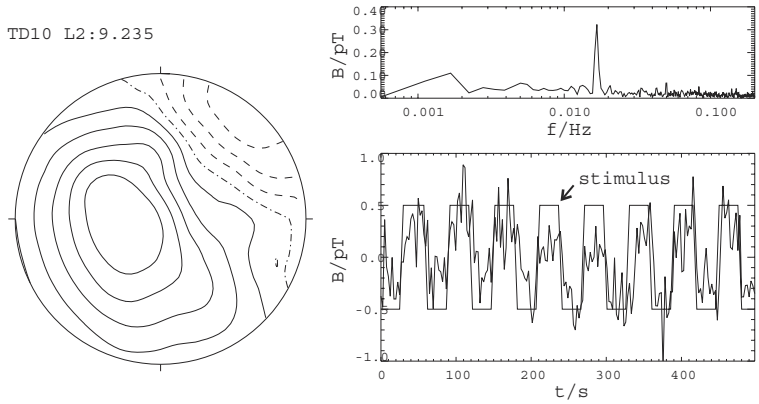


Figure 8.12 Spatial field pattern, frequency content and time course of the component TD10. Note that the extracted component follows well the stimulus. From Wübbeler et al. (2000).

We do not expect that the cortical response resembles completely the stimulus. Yet, computing the correlation coefficient between the "on/off" stimulus and the time courses of the components provides a useful measure to evaluate and compare the performance of different separation algorithms. Applying three different BSS algorithms, from sections 8.3 and 8.4, we find that, in this special application, only the TDSEP algorithm is able to recover a signal that is highly correlated to the stimulus, while FastICA and JADE yield much lower correlation coefficients (Wübbeler et al., 1998).

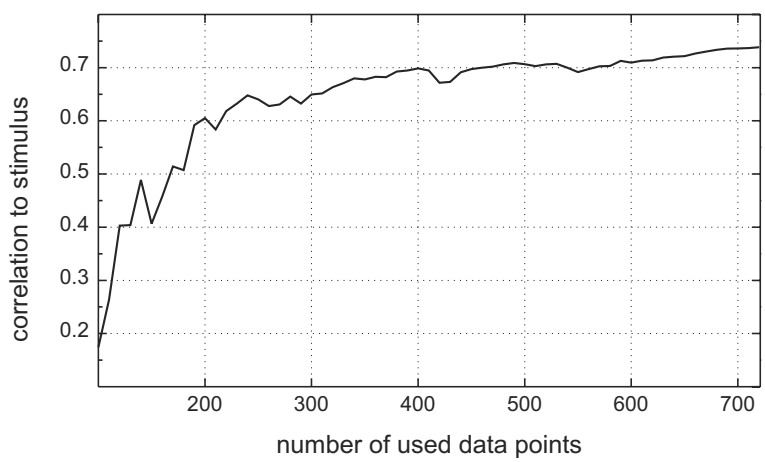


Figure 8.13 Correlation coefficient between stimulus and the best matching component versus number of samples used for TDSEP applied on the full 49-dimensional data set. From Wübbeler et al. (2000).

There might be a number of reasons for this finding. On one hand the limited number of sample points is a serious problem for algorithms based on higher-order statistics, as they have to estimate a larger amount of parameters from the same amount of data. On the other hand, the low signal-to-noise ratio is problematic as well and makes the distinction between different sources solely relying on the probability density very difficult. Furthermore we note a number of outliers in Fig. 8.10 that may harm the estimation of higher-order moments. Unfortunately simply removing potential outliers did not improve the results, as one might erroneously remove also data points which are important for a proper estimate of the higher-order statistics.

One might argue that our comparison in this specific context is unfair, as DC signals contain by definition a strong temporal correlation and may have a Gaussian distribution. However, the extracted component (TD10) from which we believe that it corresponds to interesting brain activity has a clear non-Gaussian structure ($kurtosis = -0.6$).

To investigate the effects of small sample size, Fig.8.13 shows the dependency of the separation result for TDSEP as a function of the sample size. Already for 300 samples we observe an enhanced correlation.

8.10 Concluding Remarks

What makes blind source separation an appealing method for the analysis of neurobiological data, is the reduced amount of prior assumptions required for the identification of underlying interesting features in the data. The results attained with such methods seem nevertheless to agree with more complex physiologically based ones.

In this chapter we have shown examples of BSS in the analysis of biomagnetic brain signals. A special emphasis has been given to the validation of the ICA model for EEG and MEG data. Some limitations of the BSS model we discussed as well.

ICA has shown to be able to differentiate between somatosensory and auditory brain responses in the case of combined auditory and vibrotactile stimulation. In addition, the independent components, found with no other modeling assumption than their statistical independence, exhibit field patterns that agree with the conventional current dipole models. The equivalent current dipoles corresponding to the independent components are located in brain regions expected to be activated by the respective stimuli.

Furthermore, we have used the speed and deflationary characteristics of the FastICA algorithm to derive an iterative algorithm, incorporating source modeling in its search for independent decomposition of combined somatosensory and auditory evoked responses.

Generally, the modified FastICA algorithm achieved better separation abilities, while keeping the very high agreement with the physiological plausibility of its independent components. In particular, we have seen that a fewer number of equiv-

alent current dipoles were needed to attain a better explanation of the measured recordings than when using a traditional FastICA approach. The detection of subtle information, such as lateralization effects, were as well rendered possible with the modified algorithm which were not clearly visible in the original FastICA formulation.

Finally, by using TDSEP, a BSS method that performs decorrelation at several time-lags, it became possible to extract a faithful estimate of the DC-activation level in the auditory cortex. In contrast to earlier paradigms, which identified cortical sources of short-term (2–9s) "sustained" fields (Pantev et al., 1996) or potentials (Picton et al., 1978) by averaging at least dozens of such repeated activations, the present DC-MEG plus ICA approach allows to monitor the time course of cerebral DC-activations without any need for averaging.

In an attempt to strengthen the blind separation approach, we have started to add some existing prior knowledge into it. As we depart from the purely statistically based assumptions, we get closer to physiologically plausible decomposition of electromagnetic brain signals. On the other hand, fitting neural sources in a classical framework, may be hard if some temporal overlap is present in their activations. A well balanced use of both the model and any prior is therefore needed in order to fully exploit all the advantages of each technique. We think we may have found a good middle term in determining the most independent dipole decomposition of averaged evoked responses.

Acknowledgments

We thank the members of the NIPS 2000 workshop for inspiring discussions. Valuable discussions with J. Kohlmorgen, G. Nolte and B. Blankertz are acknowledged. R.V. was funded by EU (Marie Curie fellowship HPMF-CT-2000-00813). A.Z., K.-R.M. and J.S. were partly funded by EU (IST-1999-14190 – BLISS). G.W., B.-M. M. and G.C. were partly funded by DFG grants Cu 36/1-3. E.O. is funded by the Academy of Finland.

References

T. Back, K. Kohno, and K. Hossmann. Cortical negative DC deflections following middle cerebral artery occlusion and KCl-induced spreading depression: effect on blood flow, tissue oxygenation and electroencephalogram. *J. Cereb. Blood Flow Metab.*, 14(1), 12-19, 1994.

A. K. Barros, R. Vigário, V. Jousmäki, and N. Ohnishi. Extraction of periodic signals from multi-channel bioelectrical measurements. *IEEE Trans. Biomed. Eng.*, 47(5), 583-588, 2000.

A. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7, 1129-1159, 1995.

A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on SP*, 45(2), 434-44, 1997.

S. Blanco, H. Garcia, R. Q. Quiroga, L. Romanelli, and O. A. Rosso. Stationarity of the EEG series. *IEEE Engineering in Medicine and Biology*, 395-399, 1995.

J.-F. Cardoso and P. Comon. Independent component analysis, a survey of some algebraic methods. In *Proc. ISCAS'96*, 93-96, 1996.

J.-F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *IEE Proceedings-F*, 140, 362-370. 1993.

Q. Chen, M. Chopp, H. Chen, and N. Tepley. Magnetoencephalography of focal ischemia in rats. *Stroke*, 23(9), 1299-1303, 1992.

A. Gardner-Medwin, N. Tepley, G. Barkley, J. Moran, A. Nagel-Leiby, R. Simkins, and K. Welch. Magnetic fields associated with spreading depression in anaesthetized rabbits. *Brain Res.*, 540(1-2), 153-158, 1991.

M. Hämäläinen, R. Hari, R. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2), 413-497, 1993.

R. Hari, K. Aittoniemi, M. Jarvinen, T. Katila, and T. Varpula. Auditory evoked transient and sustained magnetic fields of the human brain. localization of neural generators. *Exp. Brain Res.*, 40(2), 237-237, 1980.

G. Hinton and D. van Camp. Keeping neural networks simple by minimizing the description length of the weights. In *Proc. of the 6th Ann. ACM Conf. on Computational Learning Theory*, 5-13, Santa Cruz, California, 1993.

K. Hotary, R. Nuccitelli, and K. Robinson. A computerized 2-dimensional vibrating probe for mapping extracellular current patterns. *J. Neurosci. Meth.*, 43(1), 55-67, 1992.

A. Hyvärinen. Independent component analysis in the presence of gaussian noise by maximizing joint likelihood. *Neurocomputing*, 22, 49-67, 1998.

A. Hyvärinen. Gaussian moments for noisy independent component analysis. *IEEE Signal Processing Letters*, 6(6):145-147, 1999.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent component analysis.* Wiley-Interscience, 2001.

A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9, 1483-1492, 1997.

V. Jousmäki and R. Hari. Somatosensory evoked fields to large-area vibrotactile stimuli. *Clinical Neurophysiology*, 110:905-909, 1999.

E. Kaukoranta, M. Hämäläinen, J. Sarvas, and R. Hari. Mixed and sensory nerve stimulations activate different cytoarchitectonic areas in the human primary somatosensory cortex SI. *Exp. Brain Res.*, 63(1), 60-66, 1986.

M. Kawanabe and N. Murata. Independent component analysis in the presence of gaussian noise. Technical report, METR 2000-02 Department of Mathematical Engineering and Information Physics, University of Tokyo, 2000.

K. Knuth. Bayesian source separation and localization. In A. Mohammad-Djafari (ed), *Proc. of SPIE: Bayesian Inference for Inverse Problems*, 3459, 147-158, 1998.

H. Lappalainen and J. Miskin. Ensemble learning. In Girolami, M., editor, *Advances in Independent Component Analysis*, 75-92. Springer, Berlin, 2000.

B.-M. Mackert, G. Wübbeler, M. Burghoff, P. Marx, L. Trahms, and G. Curio. Non-invasive long-term recordings of cortical 'direct current' (DC-) activity in humans using magnetoencephalography. *Neuroscience Letters*, 273(3), 159-162, 1999.

S. Makeig, T.-P. Jung, A. Bell, D. Ghahremani, and T. J. Sejnowski. Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci. USA*, 94, 10979-10984, 1997.

F. Meinecke, A. Ziehe, M. Kawanabe, and K.-R. Müller. Estimating the reliability of ica projections. In Dietterich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems*, volume 14. The MIT Press, Cambridge, MA, 2002.

L. Molgedey and H. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 72(23), 3634-3637, 1994.

J. Mosher, P. Lewis, and R. Leahy. Multidipole modelling and localization from spatio-temporal MEG data. *IEEE Trans. Biomed. Eng.*, 39, 541-557, 1992.

K.-R. Müller, P. Philips, and A. Ziehe. JADE$_{TD}$: combining higher-order statistics and temporal information for blind source separation (with noise). In *Proc. Int. Workshop on Independent Component Analysis and Blind Separation of Signals (ICA'99)*, Aussois, 1999.

K.-R. Müller, J. Kohlmorgen, A. Ziehe, and B. Blankertz. Decomposition algorithms for analysing brain signals. In Haykin, S. and Principe, J., editors, *IEEE Symposium 2000 on adaptive Systems for Signal Processing, Communications and Control*, 105-110. IEEE Publishing, Piscataway, NJ, 2000.

K.-R. Müller, P. Philips, and A. Ziehe. $JADE_{TD}$: Combining higher-order statistics and temporal information for blind source separation (with noise). In *Proc. Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA'99)*, 87-92, Aussois, 1999.

N. Murata, K.-R. Müller, A. Ziehe, and S.-I. Amari. Adaptive on-line learning in changing environments. In Mozer, M. C., Jordan, M. I., and Petsche, T., editors, *Advances in Neural Information Processing Systems*, 9, 599-607. The MIT Press, 1997.

C. Pantev, C. Eulitz, S. Hampson, B. Ross, and L.E. Roberts. The auditory evoked off response: source and comparison with the on and the sustained responses. *Ear & Hearing*, 17(3), 255-265, 1996.

T. Picton, D. Woods, and G. Proulx. Human auditory sustained potentials: part I and II. *Electroencephalography and clinical Neurophysiology*, 45(2), 186-210, 1978.

J. Särelä, H. Valpola, R. Vigário, and E. Oja. Dynamical factor analysis of rhythmic magnetoencephalographic activity. In *Proceedings of the third international workshop on independent component analysis and blind signal separation, ICA'01*, San Diego, CA. accepted??.

J. Särelä and R. Vigário. The Problem of Overlearning in High-Order ICA Approaches: analysis and solutions. In *6th International Work-Conference on Artificial Neural Networks, (IWANN'2001)*, Granada, Spain.

M. Scherg, and D. von Cramon. Two bilateral sources of the late aep as identified by a spatio-temporal dipole model. *Electroenceph. clin. Neurophysiol.*, 62, 32-44, 1985.

B. Schölkopf, A. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10, 1299-1319, 1998.

P. Stys, B. Ransom, and S. Waxman. Compound action potential of nerve recorded by suction electrode: a theoretical and experimental analysis. *Brain Res.*, 546(1), 18-32, 1991.

L. Tong, R. Liu, V. C. Soon, and Y. Huang. Inderterminacy and identifiability of blind identification. *IEEE Trans. on Circuits and Systems*, 38, 499-509, 1991.

R. Vigário. *Independent Component Approach to the Analysis of EEG and MEG Recordings.* PhD thesis, Helsinki University of Technology, 1999.

R. Vigário and E. Oja. Independence: a new criterion for the analysis of the electromagnetic fields in the global brain? *Neural Networks*, 13, 891-907, 2000.

R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja. Independent component approach to the analysis of EEG and MEG recordings. *IEEE Trans. Biomed. Eng.*, 47, 589-593, 2000.

R. Vigário, J. Särelä, V. Jousmäki, and E. Oja. Independent component analysis in decomposition of auditory and somatosensory evoked fields. In *Proc. Int. Workshop on Independent Component Analysis and Blind Separation of Signals (ICA'99)*, Aussois, 1999.

R. Vigário, J. Särelä, and E. Oja. Independent component analysis in wave decomposition of auditory evoked fields. In *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)*, Skövde, 1998.

R. Vigário. Dipole modeling in FastICA decomposition of evoked responses. In *Proc. 2nd Int. Workshop on Independent Component Analysis and Blind Separation of Signals (ICA'00)*, Helsinki, Finland, 2000.

H.-C. Wu and J. Principe. Simultaneous diagonalization in the frequency domain (SDIF) for source separation. In *Proc. Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA'99)*, 245-250, Aussois, 1999.

G. Wübbeler, J. Mackert, F. Armbrust, M. Burghoff, B.-M. Mackert, K.-D. Wolff, J. Ramsbacher, G. Curio, and L. Trahms. SQUID measurements of human nerve and muscle near-DC injury-currents using a mechanical modulation of the source position. *Applied Superconductivity*, 6(10-12), 559-565, 1998.

G. Wübbeler, A. Ziehe, B.-M. Mackert, K.-R. Müller, L. Trahms, and G. Curio. Independent component analysis of noninvasively recorded cortical magnetic DC-fields in humans. *IEEE Trans Biomed Eng.*, 47(5), 594-599, 2000.

A. Ziehe, K.-R. Müller, G. Nolte, B.-M. Mackert, and G. Curio. Artifact reduction in magnetoneurography based on time-delayed second order correlations. Technical Report 31, GMD–Forschungszentrum Informationstechnik GmbH, 1998.

A. Ziehe, and K.-R. Müller. TDSEP – an effective algorithm for blind separation using time structure. In *Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)*, 675-680, Skövde, 1998.

A. Ziehe, K.-R. Müller, G. Nolte, B.-M. Mackert, and G. Curio. Artifact reduction in magnetoneurography based on time-delayed second-order correlations. *IEEE Trans Biomed Eng.*, 47(1), 75-87, 2000a.

A. Ziehe, G. Nolte, G. Curio, and K.-R. Müller. Ofi: Optimal filtering algorithms for source separation. In Pajunen, P. and Karhunen, J., editors, *Proc. of ICA 2000*, 127-132. Helsinki University Press, 2000b.

This page intentionally left blank

III Combination EEG/MEG and fMRI

This page intentionally left blank

9 Having Your Voxels and Timing Them Too?

Scott Makeig
Computational Neurobiology Laboratory
The Salk Institute for Biological Studies
La Jolla CA 92037

Institute for Neural Computation
University of California San Diego
La Jolla CA 92093

Tzyy-Ping Jung
Institute for Neural Computation
University of California San Diego
La Jolla CA 92093

Terrence J. Sejnowski
Computational Neurobiology Laboratory
The Salk Institute for Biological Studies
La Jolla CA 92037

Institute for Neural Computation
University of California San Diego
La Jolla CA 92093

Howard Hughes Medical Institute

The two major noninvasive functional human brain imaging modalities developed during the last part of the twentieth century, high-density scalp EEG (electroencephalogram) and fMRI (functional magnetic resonance imaging), appear from a logical viewpoint to be largely complementary. BOLD (blood oxygen level difference) signals can have a spatial resolution of less than 1 cm^3, meaning time series of blood oxygenation level differences can be recorded from many more than 10,000 spatially and structurally identifiable brain regions (voxels). However, changes in

blood oxygenation are inherently slow, much slower than the firing of individual neurons (circa 1 ms) or the frequencies at which neural networks tend to synchronize (1-200 Hz or more). EEG signals, on the other hand, can be recorded at sampling rates of a kHz or more per channel, and can thus easily and accurately record cortical potentials throughout their frequency range, if they reflect a sufficient amount or density of synchronous activity within cortex that the summed local fields reach the scalp without canceling one another. It therefore seems easy to argue, as many researchers have, that by concurrently recording EEG and fMRI BOLD signals, researchers could acquire functional brain activity data with both high spatial and high temporal resolution.

9.1 Introduction

Certainly there are formidable technical problems that researchers wishing to make such concurrent recordings must overcome. Any ferromagnetic metal inside the scanner can be dangerous for the subject, and will certainly cause unacceptable loss in the BOLD image. Then too, even very small movements of the electrodes or cables in a very strong magnetic field must induce artifactual currents. The RF pulses used in fMRI scanning are another serious challenge; their sharp gradients can induce volts of current in EEG leads that usual analog EEG high pass filters will convert into 'railed' amplifier signals. The slightest movement of the subject's scalp, including the pulse of blood through it, will move the electrodes sufficiently to produce large ballistocardiogram signals that may obscure the underlying brain activity. Finally – or not finally, as the list of potential recording problems is long – the sharp loud noise produced by the gradient pulses several times a second may generate large auditory evoked responses or induce more general perturbations of the field dynamics in the brain of the subject.

However, even after having dealt with or overcome all these and related problems, it is important to reconsider whether promoting EEG and BOLD signals as "complementary" in space and time is not too glib a concept. EEG is, first and foremost, an index of local cortical synchrony. EEG signals, certainly, reflect synchronous dendritic and possibly glial activity within domains of cortical tissue (neuropile) much much larger than a single neuron, most probably within hypercolumn-scale or larger domains. Perhaps 'neuropile synchrony' is a better term, if one gives room for active contributions of inhibitory neural networks (with their electrotonic as well as synaptic couplings) and for active contributions of non-neural glial networks. Blood oxygenation, on the other hand, is considered to index the brain response to local metabolic need in the neuropile, mediated and controlled by mechanisms whose details are not yet understood.

Since EEG and BOLD signals reflect different phenomena – spatial synchronization and total metabolic consumption, respectively – EEG and BOLD signals, even from the same patch of cortex, may be as unrelated to each other as are phase and amplitude in Fourier spectra of random signals. *That is, there is no a priori*

reason to assume they have any correlation at all! Before assuming a direct relationship between EEG and BOLD signals, we must answer the following question: *Do synchronized neuronal excitatory and inhibitory processes demand more oxygen than the same processes in a desynchronized state?* The largest normal EEG rhythms, after all, occur in deepest sleep stages, when overall brain metabolic demands (and most BOLD signals) are somewhat lower than during waking. Lack of a firm answer to this question should give us pause, and leads inevitably to the conclusion that making a *priori* assumptions about the interrelationship of EEG and BOLD signals is foolish in advance either of direct experimental evidence (from adequate concurrent EEG / fMRI studies) and/or more detailed understanding of their biophysics.

Even less well founded, in our view, are assumptions espoused by many researchers that BOLD signal increases following sensory stimuli are highly likely to indicate the brain areas responsible for generating small features (e.g., peaks) of averaged event-related potentials (ERPs) evoked by the same stimuli. We believe the very nature of the ERP may be different from that assumed by most researchers. The usual conception is the idea that sensory-evoked ERPs sum monodirectional potentials accompanying and indexing phasic stimulus-induced increases in neural firing rates observed in some neurons within limited, cytoanatomically-defined sensory cortical processing areas. In general, however, positive and negative peaks in averaged ERP waveforms may *not* index changes in total EEG energy time locked to stimulus onset, such as can be measured in the time/frequency domain by the ERSP (event-related spectral perturbation) method (Makeig, 1993). Instead, most features of averaged ERPs may be produced by event-related perturbations in the phase statistics of ongoing EEG activity (Makeig et al., 2002).

Many event-related increases in EEG spectral amplitudes, on the other hand, (as seen in ERSP plots) are not correlated with the reliable appearance of positive or negative potential peaks in the raw EEG time series. Such a correlation may occur only when the increases in EEG energy take the form of bursts that are reliably (1) time locked (e.g., peaking at the same latency, relative to event onset, across trials) and (2) phase locked (e.g., exhibiting the same phase at the same latency and central burst frequency) to the experimental events of interest. For example, following onset of briefly flashed, left-hemifield non-target squares (cueing no subject response) in a special selective attention experiment we observed a small (0.5 dB) event-related increase in EEG power near 7 Hz (figure 9.2B), whereas the ERP waveforms showed a much larger (15-dB) post-stimulus increase in spectral power which was prolonged near 10 Hz (figure 9.2A). Furthermore, the scalp topographies over which the two increases occurred were dissimilar.

Even increases in EEG spectral amplitudes (irregardless of phase) cannot be assumed to correlate with BOLD signal increases. This is clearly shown by two recent preliminary reports of negative and positive correlations, respectively between BOLD signals and EEG amplitudes in the alpha and gamma bands, respectively (Goldman et al., 2001; Logothetis et al., 2001).

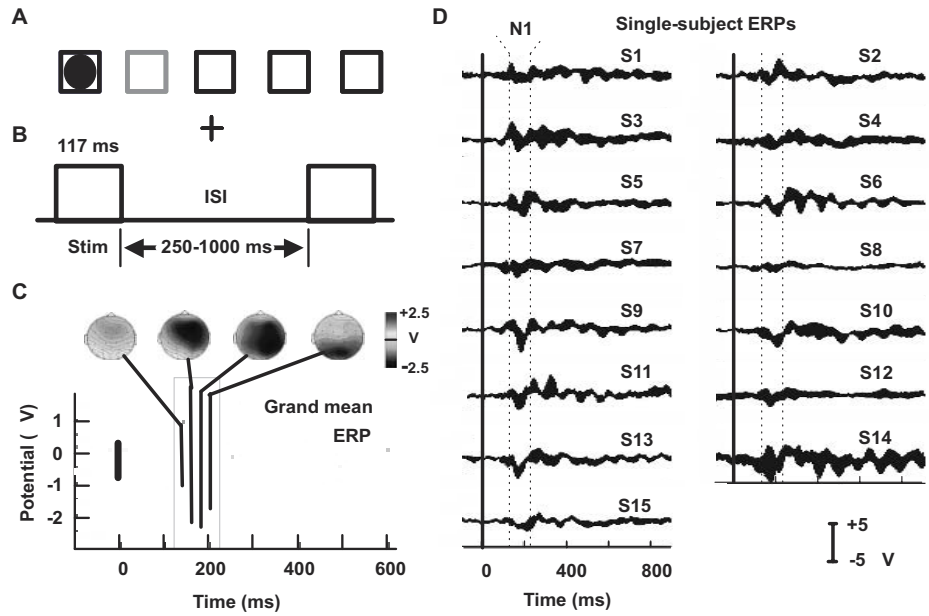


Figure 9.1 Visual Nontarget Stimulus-Evoked ERPs. **(A)** Screen display for the spatial selective attention experiment. Five 1.6 cm^2 square outlines indicating possible stimulus locations were permanently displayed 0.8 cm above a central fixation cross. In each 76-s block of trials, one outline was colored green, indicating the target location for that block of trials. Target location was evenly distributed over the five stimulus locations across 30 trial blocks per subject. **(B)** Stimulus timing. Stimuli were briefly flashed white circular discs each presented for 117 ms in a randomly selected stimulus location following a randomly selected inter-stimulus interval of 250 to 1000 ms. Subjects were asked to press a right thumb button press as quickly as possible each time a (target) stimulus appeared in the target location (green box), and to ignore (nontarget) stimuli presented in the other four boxes. **(C)** Averaged responses from 15 subjects (S1-S15). EEG data were collected from 29 scalp plus two periocular sites, referred to the right mastoid at a rate of 256 Hz/channel with an analogue band pass of 0.01 to 100 Hz. Scalp impedances were kept below 5 kΩ. After rejecting epochs containing out-of-bounds values, data were low pass filtered below 40 Hz to suppress line noise. Averaged responses to nontarget stimuli presented to the left of fixation (mean trials per subject, 922). Grand mean of 15 single-subject ERPs time locked to the brief appearance of the disk in a non-attended box to the left of subject fixation. The light blue area marks the defined N1 response interval (50 ms before and after the RMS N1 peak). The four interpolated scalp maps show the shifting scalp distribution of the averaged response during the N1 interval. Following the N1 feature, circa 10-Hz rhythmic activity appears in the evoked response. Figure 9.2B shows that this 'alpha ringing' does *not* arise from an increase in 10-Hz energy in the EEG. **(D)** Envelopes of the 15 single-subject ERPs. The solid blue response envelopes enclose the individual response traces for all 29 scalp channels. Vertical dashed lines mark the grand mean N1 interval. From Makeig et al., Science 295, 690-694, 2002. Reprinted with permission.

EEG signals are produced by spatial synchronization of electrochemical activity in cortex; ERPs by synchronization of EEG signals time locked to the occurrence of some class of experimental events. Recently, we have shown that the electroencephalographic (EEG) response to a small visual stimulus presented in an unattended location in a selective spatial attention task (figure 9.1) is better modeled by stimulus-induced phase realignment of EEG activity within domains of strong cortical synchrony that appear to generate most of the ongoing EEG (Makeig et al., 2002) (figure 9.2). These EEG domains need not be located within single functionally-defined cortical processing areas. Some evidence suggests they may extend across boundaries between such areas (Rogeul-Buser et al., 1997).

If visual ERPs are produced by stimulus-induced phase resetting of multiple ongoing EEG processes rather than by consistent evoked positive or negative potentials generated in restricted cortical areas, as strongly supported by our results and other reports as early as Sayers et al. (1974), it appears naive to assume that generators of single ERP peaks should be co-located with areas of significant event-related BOLD signal change. BOLD signal changes should reflect changes

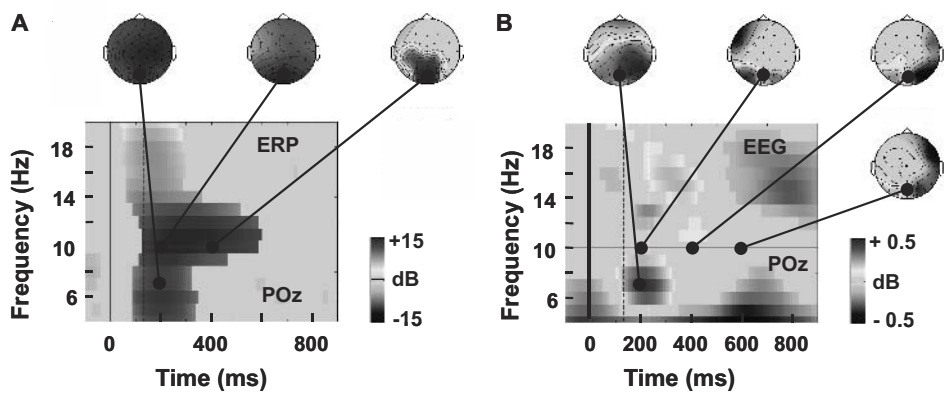


Figure 9.2 Differences between Event-related Power Spectral Changes in the ERP and EEG.. **(A)** Event-related spectral perturbation (ERSP) plot showing mean post-stimulus increases in spectral power of the single-subject ERPs, averaged across 15 subjects. Shaded areas in the time/frequency plane that show significant ($p < .02$) post-stimulus increases or decreases (see gray scale) in log spectral power in the averaged ERP waveform at a central parietal electrode site (POz) relative to mean power in the averaged 1-s pre-stimulus ERP. Topographic scalp maps show topography of the post-stimulus power increases in the ERP relative to baseline across all 29 scalp channels at three indicated points in the time/frequency plane. **(B)** Event-related spectral perturbations (ERSPs) for the unaveraged EEG at central posterior site POz following left visual field nontarget stimulus presentations. Mean of similar time/frequency transforms 15 subjects. Log spectral power at each time and frequency was normalized by subtracting mean log power in the 1-s pre-stimulus baseline. Features near (7 Hz, 200 ms) and (16 Hz, 350 ms) reflect small power increases. *Vertical dotted lines*, the N1 interval; *horizontal line*, 10 Hz. *Cartoon heads*, scalp topographies of differences in spectral amplitude, relative to baseline, at the indicated time-frequency points. From Makeig et al., Science 295, 690-694, 2002. Reprinted with permission.

in the level of neural or neuroglial activity in local cortical domains whose size is determined by the voxel size of the measurement (convolved with hemodynamic control patterns). Even if some ERP peak were generated predominantly in a single compact cortical area, it is premature to assume that a relatively slow BOLD signal increase should be triggered by a brief period of net positive or negative far-field potential, as this might well reflect a transient increase in synchronization of synaptic activity rather than an increase in its metabolic activity level.

If the time courses of averaged ERPs do not accurately reflect event-related dynamic patterns in the unaveraged EEG signals, which may in turn differ in complex ways from trial to trial, then combining averaged BOLD and EEG/ERP data collected in separate sessions is an idea built on shaky ground indeed. Here we report results of a single-subject, single-channel pilot study to test the feasibility of recording and analysis of concurrent EEG and BOLD data. Preliminary results from this study were reported earlier (Jung et al., 1999).

9.2 Methods

EEG epochs time locked to presentation of an RF pulse during an oddball/rare-target "P300" experiment at the vertex to a flashed target shape (in the fovea), from a pilot experiment conducted. The data (232 trials) were collected **during** continuous, concurrent EPI scanning using a Siemens 1.5-T scanner at a rate of 5 slices per second in a blocked paradigm consisting of three 4-minute bouts each consisting of 40-s task blocks consisting of presentations of standard circles (75%) and target squares (25%)) alternating with 40-s control blocks during which the nontarget stimuli only were presented and the subject was instructed only to continue to fixate a dot at screen center while ignoring the flashed stimuli. Stimulus SOAs ranged between 435 and 1116 ms. During task blocks the adult volunteer subject was asked to attend to differences between the visual stimuli and to push a handheld button as soon as possible after presentation of a target (square) stimulus. The subject wore earplugs to minimize sound from the scanner during scanning pulses. Behavioral responses to targets were collected via a non-metallic thumb button held in the subject's right hand. EEG was recorded from a tin electrode placed over the right central scalp and referenced to another placed over the right mastoid. The electrodes were attached to the amplifier by carbon wire (Electrocap, Inc.). The recording used an analog pass band of 0.1-30 Hz and a sampling rate of 500 Hz. The amplifier (SA Instrumentation, Inc.) was especially constructed to operate on batteries placed before and after an optical bridge, and to 'time out' during a pulse produced by the Siemens scanner from a few ms before to approximately 16 ms after the production of each RF scanning pulse. During these brief time-out periods, capacitors held the EEG signal level nearly constant to minimize switching artifacts.

9.3 EEG Artifact Removal

First, the mean noise waveform time locked to the RF pulse was removed from the data. The relative consistency of this noise is indicated by figure 9.3, which shows EEG epochs time-aligned to 100 consecutive RF pulses using a relative microvolt scale. Unfortunately, for this pilot experiment microvolt calibrations were not available. Time 0 marks the onset of the RF pulse.

Figure 9.3 shows the time-out period (with 2 sharp but relatively small remaining spike artifacts), and a larger 10-Hz artifact time locked to pulse presentation. The precise origin of the 10-Hz wave is unknown – two obvious candidates are ringing at the frequency of the scanning pulse (near 1500 Hz) aliased down to near 10 Hz by the 500-Hz EEG sampling-rate, and/or possibly, entrainment of subject alpha/mu activity by the pulse (and accompanying loud noise burst) sequence. The mean artifact was then subtracted from the EEG data surrounding each RF pulse. (Pulse onsets were recorded on a separate recording channel). Next, EEG epochs surrounded the 226 target stimulus presentations were extracted from the

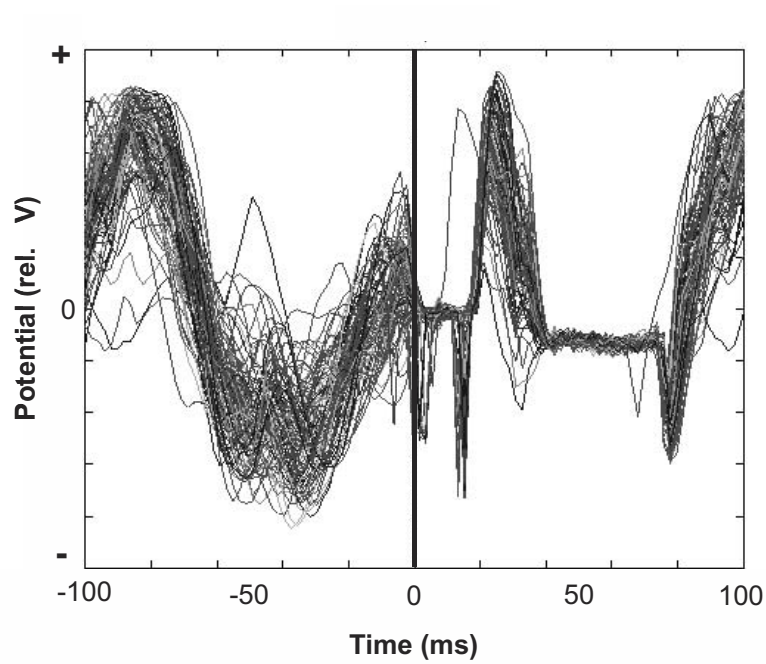


Figure 9.3 Scan-related EEG artifact. One hundred consecutive EEG epochs time locked to the onset (at time 0) of RF pulses during fMRI scanning. The ~80 msec timeout period is followed/preceded by a stereotyped noise waveform to which is added EEG variability occurring during the ~120 msec non-blanking periods. Removing the mean pulse-locked artifact from each pulse epoch allows reconstructing the variable EEG and ERP activity in single trials (see Figs. 2-3). Y-axis scaling is proportional to μV.

cleaned data and imaged using the ERP-image plotting technique (Jung et al., 1999; Makeig et al., 1999). Finally, we performed ERSP and phase resetting analysis on the cleaned data. The figures were created using an ICA/EEG toolbox for MATLAB (The Mathworks, Inc.) available for download on the Web (sccn.ucsd.edu/eeglab/).

9.4 Results

Single-trial results are summarized in figure 9.4 using the 'ERP-image' plotting format (Jung et al., 1999; Makeig et al., 1999).

The striate pattern visible in the pre-stimulus EEG data (and throughout) reflects the remaining pulse-locked artifact. However, as the delivery of stimuli was not time locked to the production of scanning pulses, the ERP-image format clearly reveals event-related potential activity in single trials (horizontal image lines) which here are sorted in ascending order of response latency (indicated by the sloping white line). The cleaned data clear show the presence of a complex pattern of evoked response activity time-locked to the button press. The trace below the ERP-image shows the average ERP of the cleaned trials.

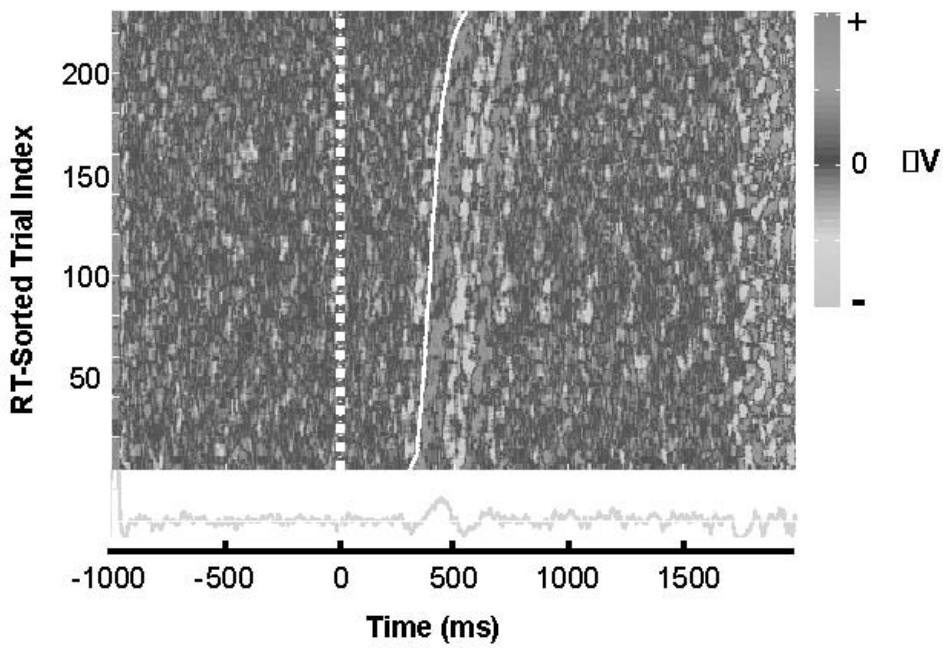


Figure 9.4 Target stimulus-aligned ERP-image plot. This plot shows the 226 single trials after removing the mean pulse-locked artifacts from each trial. The dotted white line shows the stimulus onset time. The curved solid white line shows subject response time (RT) after sorting trials by RT. Both the data and the RT curve are smoothed with a 20-trial moving average. The averaged target-locked response is shown below the image of the single-trial data.

As figure 9.4 clearly shows, the averaged ERP is smeared because of the relatively broad range of response times in different trials. The lower trace of figure 9.5 shows the response-locked ERP for the same trials. Note the ∼10-Hz activity in the ∼100 ms following the response (solid white line). This may represent the phase resetting of mu activity, as suggested by other results obtained out of the scanner. The two larger peaks of the averaged evoked response are separated by approximately 200 ms, suggesting a circa 5-Hz character.

Figure 9.6 presents time/frequency analysis of the same target epochs. The top panel shows the ERSP, which includes a strong (6-dB) mean increase in theta band power (near 5 Hz) in single trials during the evoked response. This theta power increase does not in itself produce the 5-Hz ERP features. Instead, as shown in the lower panel, the phase of theta bursts following the motor response is not random. This switch from a pre-stimulus random to a post-stimulus non-random phase distribution across trials, termed 'phase resetting,' is indexed by the significant (r=0.4) inter-trial coherence in the theta band. More exactly, the term should be 'partial phase resetting,' since the resetting is incomplete ($1 \gg r \gg 0$).

The top panel also shows that the increase in EEG following the subject response is not confined to one frequency band. Instead, a similar though smaller phasic

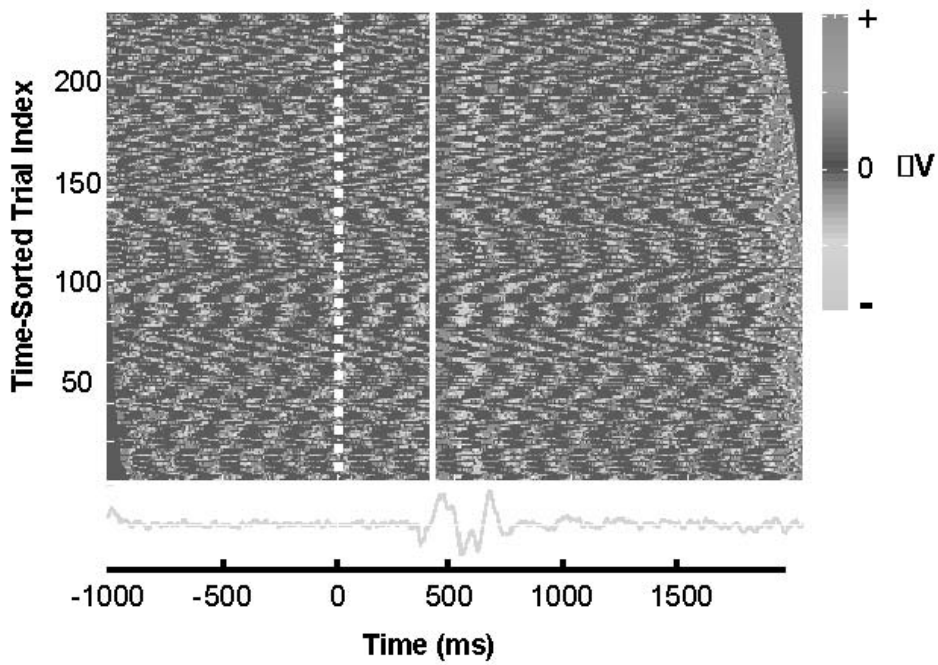


Figure 9.5 Response-aligned ERP-image plot. This plot shows the 226 single trials after removing the RF-pulse artifacts from each trial. The trials are aligned to median subject reaction time (403 ms, *solid white line*). *The dotted white line* shows the median stimulus onset time. The mean RT-aligned response is shown below the single-trial image. Notice the additional small peak (after 500 ms) in the response-locked average response (lower trace); this feature does not appear in the stimulus-locked average (figure 9.4 lower trace).

power increase is seen near 14 Hz, 18 Hz, and 38 Hz. Near 38 Hz, the augmentation begins just after stimulus onset and continues throughout the epoch, again peaking after the button press. The augmentations at these higher frequencies are not accompanied by phase resetting, since phase following the stimulus is random (ITC not significantly different from 0). However, figure 9.6 measures ITC relative to stimulus onset.

Figure 9.5 suggests significant phase resetting occurs time locked to the motor response in at least two bands (near 5 and 10 Hz). The occurrence of partial phase resetting in the theta band time locked to both the stimulus and the motor response is quite possible. Further research would be required to determine if this occurred in different trial subsets or only as a consequence of the response time distribution being concentrated in less than a 200 ms window representing one 5-Hz cycle (see figure 9.5). We are examining the relation of phase resetting to perceptual awareness and behavior in more extensive EEG-only data sets.

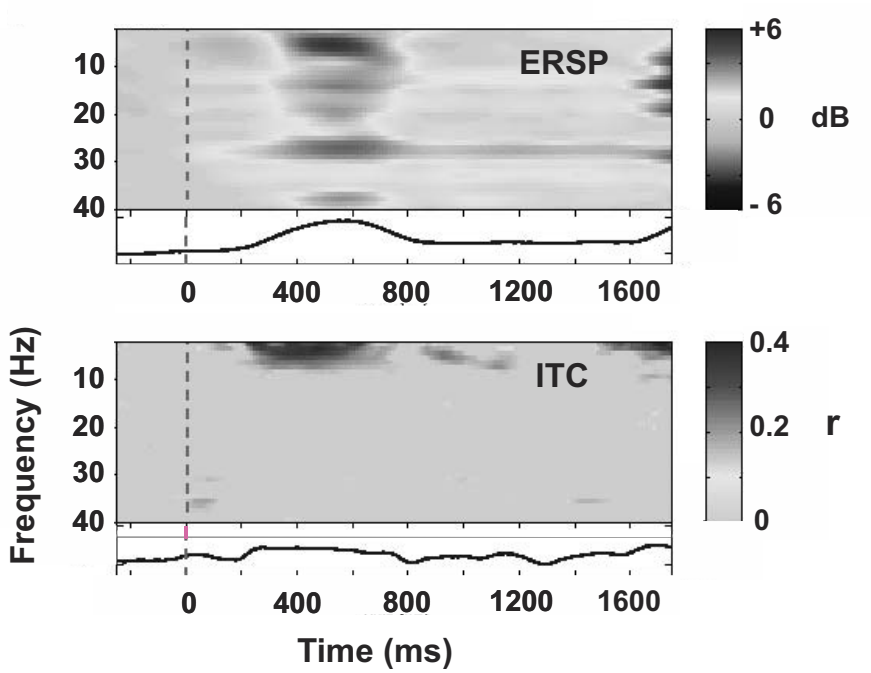


Figure 9.6 Frequency domain characterization of target event-related EEG dynamics. (*Upper panel*) Shaded areas show significant transient event-related power increases in the EEG spectrum (the event-related spectral perturbation, ERSP) during the 232 target response epochs time locked to stimulus onset. In addition to the circa 4-8 Hz theta band power increase during the period of the late evoked response (400-700 ms), there are increases near 12, 19 and 38 Hz, plus a lengthy post-stimulus increase in power near 28 Hz. (Increases after 1500 ms are probably edge artifacts). (*Lower panel*) Significant changes in inter-trial coherence (ITC) measuring changes in consistency of EEG phase-locking of the EEG at each frequency to stimulus onsets. None are evident save for the ERP-related circa 4-5 Hz peak. Thus, the 28-Hz augmentation in the upper panel represents an event-related amplitude modulation of 28-Hz activity without phase resetting.

9.5 BOLD Signal Analysis

Straightforward correlation analysis was performed on the BOLD data by correlating its time course at each brain voxel with a task design reference function obtained by convolving the alternating block design time course with a model hemodynamic response function. Results, shown in figure 9.7, included expected activations in left motor cortex at or near the expected location hand motor area. Artifacts from the electrodes were confined primarily to supra-brain areas. Further analysis of this pilot experiment data was not attempted. Currently, we are performing experiments using a continuous performance task and 72 similar scalp electrodes, with an intention to compare the spatially independent components of the BOLD signals (McKeown et al., 1998) with the dynamics of temporally independent components of the concurrently recorded EEG (Makeig et al., 1996).

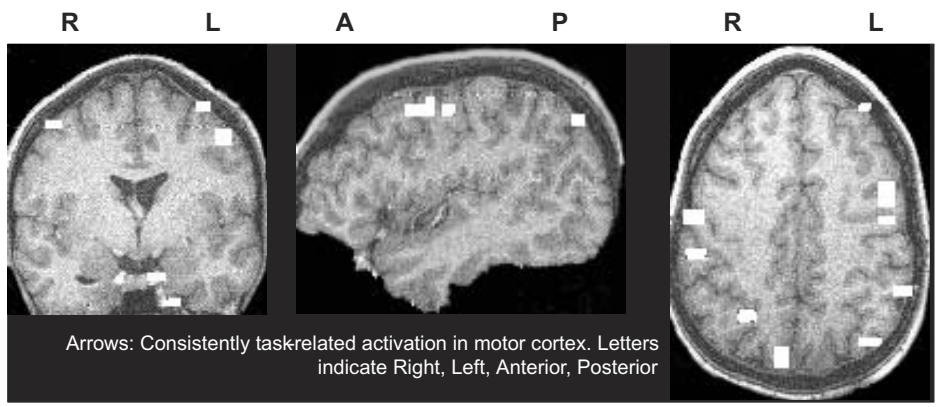


Figure 9.7 Voxels whose BOLD signals are positively correlated ($r > 0.3$) with the task block design. BOLD data recorded during concurrent EEG acquisition on a Siemens 1.5-T scanner shows BOLD activation in left motor cortex during a visual detection task requiring right thumb button presses.

9.6 Discussion

EEG signals recorded from the scalp arise through synchronous activity in cortical domains or networks. Cortical BOLD signals, on the other hand, are believed to index the brain response to total metabolic demand. In theory, these may be as uncoupled as phase and amplitude are in noise signals. As figure 9.7 also depicts schematically, single pyramidal cells in cortex can only fire upon receiving sufficient synchronous excitatory input. It is now becoming clear that synchronization of activity across and between cortical areas can effectively bias or modulate the firing of single neurons. Thus, information transfer in cortex is also controlled in part by the network synchronies that give rise to EEG. If changes in EEG power

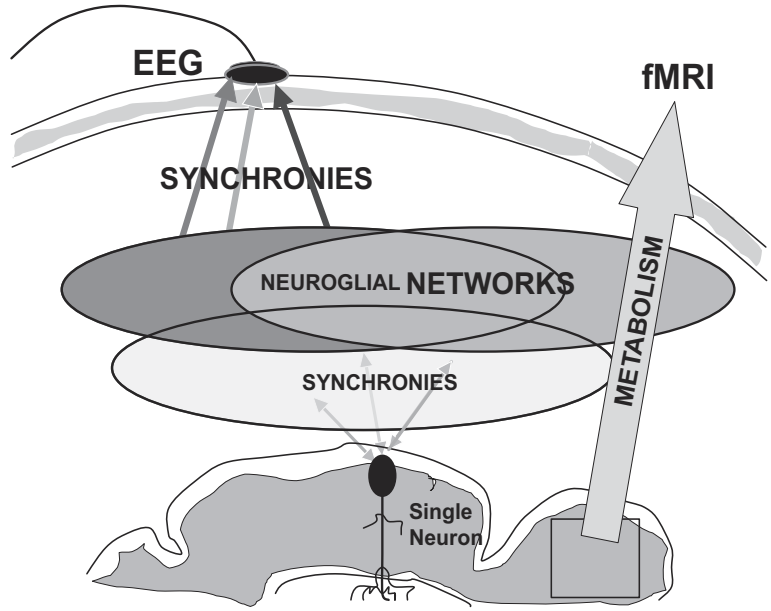


Figure 9.8 EEG and fMRI BOLD data measure different aspects of cortical activity. Whereas the BOLD signal is thought to measure the brain hemodynamic response to local metabolic need, the scalp EEG sums volume-conducted potentials generated in cortical domains/networks (of unknown size, shape and density) across which extracellular potential varies with sufficient synchronicity. Synchronous activity, as reflected in far-field potentials, can also modulate firing of individual neurons. Changes in the degree of synchronization of activity in a cortical area may not mirror changes in total metabolic consumption; thus changes in EEG power need not be correlated with BOLD signal changes.

and in BOLD signal strength are significantly related, we suggest the nature of their links cannot be guessed in advance, may prove complex, and may eventually be understood to arise through biophysical mechanisms whose details are not yet discovered.

Clearly, the true test of these predictions will come from sufficient analysis of a wide range of recordings of concurrent EEG and BOLD signals. The pilot data we have shown here indicate that detailed analysis of the dynamics of concurrently recorded EEG and BOLD signals is feasible, even to the extent of single-trial analysis of human cognitive ERP features. However, we have argued that the most intimate relationship between BOLD and EEG signals is not likely to be between BOLD differences and sources of peaks in ERP scalp waveforms, as other have suggested, but between BOLD signals and changes in power and/or other whole-signal features of the scalp EEG. Early results in this direction are promising. Logothetis and colleagues (2001) have reported preliminary results from a relatively few cells in a small cortical area of anesthetized monkeys that indicate that neural firing rate may not be as *positive* a correlate of BOLD signals as is changes in

the power of local field potentials in the gamma band (above 30 Hz). Goldman and colleagues (2001), however, have reported that BOLD signal levels within discrete cm-scale domains of posterior cortex in humans are *negatively* correlated with alpha band EEG power on the posterior scalp, while in other places (e.g., in insula, thalamus) the same correlation may be positive. Clearly, we expect that concurrent EEG and fMRI studies will prove important for the development of cognitive neuroscience, and possibly for neuroscience in general (figure 9.8).

Acknowledgments

We acknowledge the essential contributions of Jeanne Townsend, Marissa Westerfield, Rick Buxton and Eric Wong of UCSD and of Gary Priebe of SA Instirumentation, Inc. in the collection of the data, as well as the financial support of the National Institutes of Health and The Swartz Foundation.

References

R. Goldman, J. Stern, J. Engel, and M. Cohen. Tomographic mapping of alpha rhythm using simultaneous EEG/fMRI. Human Brain Mapping, Brighton, UK, 2001.

T-P Jung, S. Makeig, J. Townsend, M. Westerfield, B. Hicks, E. Courchesne, and T.J. Sejnowski. Single-trial ERPs during continuous fMRI scanning. Society for Neuroscience Abstracts, 1999.

N. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150-157, 2001.

M. J. McKeown, S. Makeig, G. G. Brown, T.-P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski. Analysis of fMRI data by decomposition into independent spatial components. Human Brain Mapping 6, 1-31, 1998.

S. Makeig. Auditory Event-related dynamics of the EEG spectrum and dffects of exposure to tones. Electroencephalogr Clin. Neurophysiol. 86, 283-293, 1993.

S. Makeig, A. J. Bell, T-P. Jung, and T. J. Sejnowski. Independent component analysis of electroencephalographic data. Advances Neural Information Processing Systems 8, 145-151, 1996.

S. Makeig, M. Westerfield, T-P. Jung, J. Townsend, E. Courchesne, and T. J. Sejnowski. Dynamic brain sources of visually evoked responses. Science 295, 690-694, 2002.

A.Rogeul-Buser and P. Buser. Rhythms in the alpha band in cats and their behavioural correlates. Int. J. Psychophysiology 26, 191-203, 1997.

B. M. Sayers, H. A. Beagley, and W. R. Henshall. The mechanism of auditory evoked EEG responses. Nature 247, 481-483, 1974.

This page intentionally left blank

10 Recording of Evoked Potentials during Functional MRI

Frithjof Kruggel and Christoph S. Herrmann
Max-Planck-Institute of Cognitive Neuroscience
Stephanstraße 1, D-04103 Leipzig, Germany
{kruggel, herrmann}@cns.mpg.de

Recording event-related potentials (ERP) and observing the hemodynamic response by functional MRI (fMRI) scanning are two complementary techniques for studying brain functioning. A combination of both methods promises to reveal more detail about the brain processes employed in a cognitive task. However, the interfering measurement conditions require the solution of a number of delicate technical details: influences of the electrode/amplifier set-up on MRI scanning and of the scanning process on the recording of electrophysiological signals are reviewed. Implications and limitations of conducting combined ERP/fMRI experiments using higher-level cognitive stimuli are discussed on the basis of two example studies.

10.1 Introduction

Electric potentials and the hemodynamic response of the vascular system are measurable correlates of the brain's neuronal activation. The first effect, measured here by event-related potentials (ERP), is a direct consequence of synchrony in the electrical activity of neurons, and allow the observation of aspects of the underlying cognitive brain process on a millisecond timescale. The second effect, measured here by functional magnetic resonance imaging (fMRI) is only indirectly linked to the energy consumption of the neuronal population and takes place on a timescale which is of the order of seconds. However, recent developments in experimental techniques and data analysis have shown that hemodynamic responses are indeed modulated by the experimental stimulation and carry information about the underlying processes at least on a 100 ms timescale (Buckner, 1998; Clark et al., 1998; Kruggel and von Cramon, 1999; Richter et al., 1998).

The localization of an activation by ERP source analysis provides a good temporal resolution but suffers from poor spatial resolution and the theoretical problem of providing only inexact solutions. Here, fMRI is better able to localize brain activations at a high spatial resolution. A combination of both techniques is a very attractive aim in neuroscience, and a number of research groups have taken up the challenge. Most studies so far were performed as separate experiments (i.e. ERP and fMRI recordings at different times), and results were registered and combined by data processing (e.g., Liu et al. (1998); Martinez et al. (1999); Opitz et al. (1999); B.R. Rosen (1998); Woldorff et al. (1999)). Especially for cognitive stimuli, it is impossible to control whether a subject performs in the same manner in both experiments: a response may habituate due to stimulus repetition; solution strategies may change. Thus, performing a single experiment, i.e., recording ERPs during fMRI, offers the advantage of observing responses to an unique event. However, constraints imposed by a combined measurement will likely always lead to results which are inferior in quality to separate measurements. It is an open issue, which way of conducting experiments yields more valuable results. For observing "unrepeatable events" such as epileptic fits or sleep stages in the spontaneous electroencephalogram (EEG), combined measurements are a pre-requisite.

However, recording EEG during fMRI scanning reveals a number of delicate technical problems, which will briefly be summarized in the following section. A typical experimental set-up for conducting ERP/fMRI measurements is described, and implications for experimental design are discussed on the basis of two studies.

10.2 Technical Considerations

The first report on recording EEG during fMRI scanning by Ives et al. (1993) already mentioned the three major problems confronting combined measurements:

- The strong and rapidly changing radio-frequency (RF) fields coupled with the large static magnetic field of the MR scanner may introduce significant current flow within the electrodes and wires, potentially leading to RF burns on the contact surface of the electrodes on the skin. Lemieux et al. (1997) identified RF-induced electromotive forces as the most important potential hazard. Placing a 15 kΩ current-limiting resistor close to the electrode is sufficient to ensure patient safety. Heating of the electrode gel was considered negligible.
- The EEG equiment may interfere with the imaging quality of the scanner. Krakow et al. (2000) investigated the effect of individual components of the EEG recording equipment on image quality. Electrodes and wires lead to local signal drop outs and to geometric distortions due to magnetic susceptibilty differences (e.g., at the electrode-skin contact) and the presence of Eddy currents in EEG electrode assemblies. Secondly, the electromagnetic noise emitted by the EEG recording unit degradates the image signal-to-noise ratio (SNR). Carbon electrodes with carbon wires introduced the smallest artifacts. Due to their delicate handling, these authors

suggest using gold electrodes and cermet film resistors as a second best choice, which limit areas of signal loss to extracortical regions. Fast switching of the EEG digitizing circuitry induces broadband signals decreasing the image SNR, so that the recording unit needs adequate shielding.

■ Various superimposed RF and gradient fields induce voltages which are much higher than the brain's response and thus interrupt electroencephalogram (EEG) acquisition (often called imaging artifact, Allen et al. (2000)). For this reason, most EEG/fMRI studies have used clustered scanning protocols (Goldman et al., 2000; Kruggel et al., 2000a), i.e. the brain region of interest is scanned rapidly at a fraction of the repetition time (TR), leaving the remaining time of the period "silent" of RF pulses for an undisturbed EEG recording. After each scan, the amplifier needs 50-100 ms to recover from saturation, which must be taken into account in the experimental design. Example: in an event-related design with a TR of 8 s, the first 2 s after a stimulus may be silent for EEG recording, then 6 s are left for fMRI scanning.

The first experimental EEG set-ups have quickly turned into commercially available MR-compatible EEG amplifiers, which incorporate provisions as discussed above. Most studies so far used only a small number of electrodes (up to 20), while conventional ERP experiments are run with 64 or 128 electrodes. Using a higher number of electrodes is necessary to cover a higher portion of the scalp (and the brain underneath) and to achieve a better spatial resolution of the recorded biosignals. However, when using a high electrode and wire density, influences on fMRI acquisiton are more likely.

In addition, a pulse-synchronous artifact is encountered in the recorded EEG, which reaches amplitudes of 100-500 μV and thus hides the real EEG signal. Head movements induced by heart action (the so-called cardio-ballistic effect, Ives et al. (1993); Müri et al. (1998)) lead to small movements of the electrodes and wires in the magnetic field, and thus induce a voltage in the wires. This effect scales with magnetic field strength and is in fact the dominating signal component at 3 T (Kruggel et al., 2000a). Twisting wires pairwise and careful subject fixation minimize this artifact. However, post-hoc correction methods are required to eliminate remaining artifact components from the signal (Allen et al., 1998; Goldman et al., 2000; Müri et al., 1998; Kruggel et al., 2000a).

In summary, to obtain a similar ERP quality under the interfering measuring conditions of combined experiments, we estimate that 2-3 times the number of trials must be conducted (Kruggel et al., 2001).

10.3 Experimental Design

In principle, both blocked and event-related experimental designs are feasible for a combined ERP/fMRI experiment, and any fMRI experiment may be enhanced by ERP recording. Some of the technical issues discussed above lead to constraints

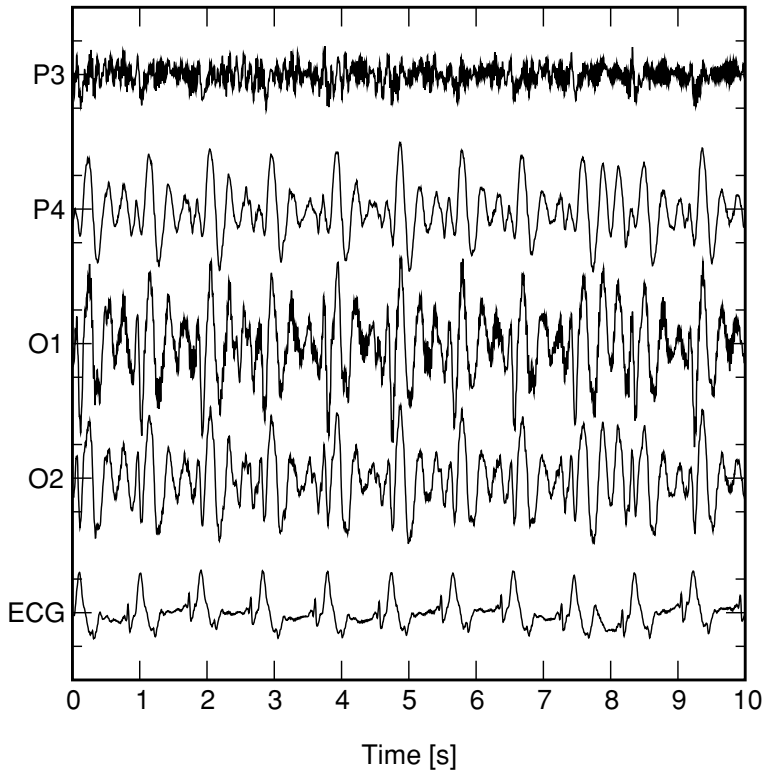


Figure 10.1 Signal recorded *inside* the scanner at sample scalp positions and corresponding ECG (tick marks on EEG traces: ± 200 μV, on ECG trace: ± 1 mV). This biosignal is about five times in magnitude compared to recordings outside the scanner.

in the design of an experiment. As a consequence of the imaging artifact, a silent period is required for EEG recording, thus, there is a tradeoff between the length of this period and the number of slices acquired within a TR. In both experiments described below, we limited scanning to a sparse set of slices. Thus, slice positions must be chosen carefully to cover the interesting brain regions. It is useful to conduct a fMRI pre-study to define the components of the network under study. A consequence of the interfering measurement conditions is a low SNR of the obtained ERPs, so that typically 100-250 repetitions per stimulus class are necessary to obtain an ERP quality which is comparable to a separate recording. Thus, even a simple factorial design requires 500-1000 trials. Given the rather uncomfortable and confined situation for a subject in the MR tunnel, the duration of an fMRI experiment (excluding preparatory scans) is limited to less than 45 min. Thus, the trial length may not exceed 2.5-5 s, which leads to strongly overlapping hemodynamic responses. Burock et al. (1998) demonstrated that disentangling responses from rapid presentation rates is possible when using a randomly varied trial length. Nevertheless, short trial lengths may not be feasible

for some cognitive stimulus material, such as used in auditory speech perception or recognition memory tasks. Auditory experiments may also be disturbed by the scanning noise, which may also lead to lateralization of brain responses (Herrmann et al., 2000b).

10.4 Experiments

We now describe two sample studies to demonstrate the feasibility of conducting combined ERP–fMRI experiments. The first experiment aims at recording a visual evoked potential (VEP) using an alternating checkerboard stimulus in a blocked design. Details of the experimental set-up and data analysis will be discussed for this simple experiment. As a typical example for a functional study in cognitive neuroscience, the second experiment employed a well-studied visual oddball task using illusory figures (Herrmann et al., 1999, 2000a). This design incorporates 4 conditions with 225 repetitions each and a randomly varied trial length of 2–3.5 s. Methods for *post-hoc* artifact correction are described, as well as a non-linear model for disentangling the strongly overlapping hemodynamic responses.

10.4.1 Recording a VEP

Five healthy persons took part in this study (3 female, 2 male, mean age 23.6 years, range 21-27 years). Conventional plastic-coated Ag/AgCl electrodes with iron-free copper leads of 60 cm length were fixed on the subject's scalp by a stretchable plastic cap. Electrodes were mounted at all positions of the international 10/20 system except Pz, where leads left the cap. The reference electrode was placed close to the nasion on the forehead. Wires were twisted pairwise and led through a flexible silicon tube to the EEG amplifier. In order to minimize movements, the subject's head was restrained using cushions. Cables and amplifier were fixed to the gantry by tape and weighed down by rice bags.

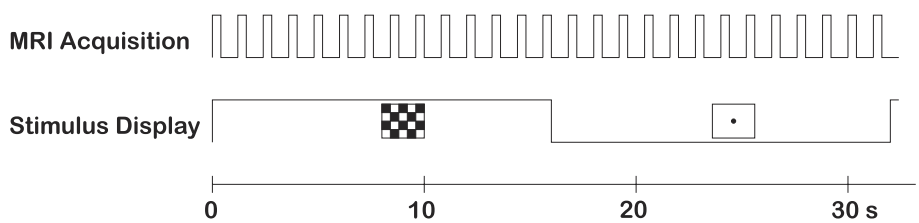


Figure 10.2 Scheme of experiment 1: Sixteen seconds of alternating checkerboard display were followed by a 16 s display of a fixation cross. A clustered acquisition of 3 MRI slices during a period of 200 ms was performed at a repetition time of 1333 ms, leaving a 1133 ms period for EEG acquisition.

To elicit visual evoked potentials, a black/white checkerboard pattern of 16x16 patches (full field visual angle 11.5 degrees, 42 arc minutes per pattern) was inverted in intervals of 550 ms (a trial) (Celesia and Brigell, 1999). 16 s checkboard stimulation were followed by a 16 s display of a fixation point. 16 blocks were conducted (corresponding to a total of 310 trials).

10.4.1.1 Data Recording

A commercially available MR-compatible system (Schwarzer, München, Germany) was used for EEG recording. The battery-powered amplifier located in the scanner tunnel was connected via a 20 m fiber optic link to a standard PC equipped with a digital signal processor (DSP) board in the MR console room. The DSP board received trigger input from the stimulation PC which was recorded with the biosignals. The amplification factor of the system was 10.000 x, with a bandwidth of 0.073-70 Hz. Biosignals were sampled at 250 Hz using an unipolar recording with Fz as reference. Functional imaging was performed using a Bruker Medspec 30/100 3.0T MR system. An in-house EPI implementation (TE 30 ms, TR 1333) was used to acquire three slices (19.2 cm FOV, 64x64 matrix, 5 mm thickness, 2 mm gap) centered along the calcarine fissure. The clustered image acquisition time was 200 ms, leaving a 1133 ms period for EEG acquisition.

10.4.1.2 EEG Data Analysis

Recorded EEG data were processed offline in a series of steps. *Slow-frequency components* of the signal were removed by a Hamming-weighted 0.8 Hz high-pass filter. The imaging artifact was detected in the summed signal. If the slope of this signal exceeded a threshold of 25 $\mu V/ms$, an interval of the following 200 ms was marked for exclusion. Then, the *cardio-ballistic artifact* was corrected: first, the length of all cardiac cycles in the recording were detected from the ECG trace using the first peak of the autocorrelated signal. A tracewise model of the cardio-ballistic artifact was then computed by cutting the trace into sections corresponding to the actual length of the cardiac cycle, interpolating each section to a length of 1000 points, and subsequent averaging. This artifact model was adapted to the original signal by varying an amplitude factor, a offset potential and a temporal shift. The adapted artifact model was subtracted from the signal, which was low-pass filtered using a cut-off frequency of 30 Hz to yield the corrected EEG. Finally, all trials were averaged, with marked intervals excluded (see Fig. 10.4).

10.4.1.3 fMRI Data Analysis

Analysis of fMRI data consisted of a series of steps: *Subject movements* were corrected in 2D (two translational and one rotational parameter) within and between both scans (Friston et al., 1996). *Baseline filtering* was achieved by estimating the baseline using low-pass filtering in the temporal domain (cut-off

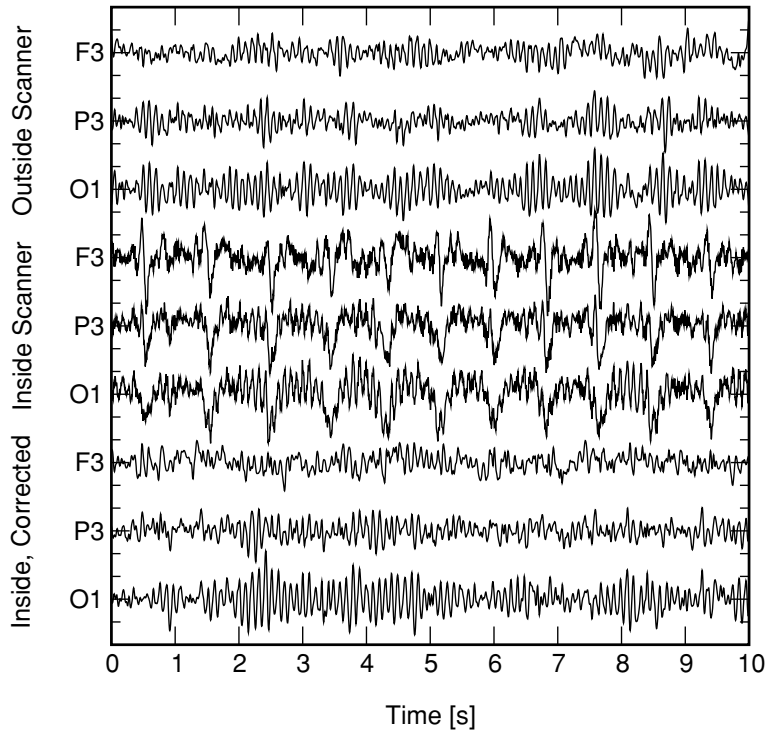


Figure 10.3 Spontaneous EEG *outside* the scanner (top, tick marks: ± 50 μV), *inside* the scanner (middle, tick marks: ± 200 μV), and EEG corrected for the cardio-ballistic artifact (below, tick marks: ± 50 μV). Trace labels correspond to electrode locations.

0.05 Hz) and subtraction of the result from the data (Kruggel et al., 1999). *System and physiological noise* were partially removed by low-pass filtering in the temporal domain (cut-off 0.2 Hz) (Kruggel et al., 1999). *Functional activation* was detected by voxelwise univariate regression analysis using a box-car waveform shifted by 5 s to match the lag of the hemodynamic response (Kruggel and von Cramon, 1999; Worsley and Friston, 1996). F-scores were converted into z-scores, thresholded ($z > 8$) and activated blobs were assessed for significance on the basis of their spatial extent (Friston et al., 1994). For graphical display, significantly activated brain areas were color-coded and overlaid onto T_1-weighted anatomical scans obtained at the same positions as the functional data (see Fig. 10.4). As expected, the striate cortex was activated by the stimulation. Components and latencies of the evoked potentials are in accordance with published data (Celesia and Brigell, 1999).

10.4.2 Recording ERPs due to Cognitive Stimuli

The second experiment employed a well-studied visual oddball task using illusory figures (Herrmann et al., 1999, 2000a). Twelve healthy persons took part in this

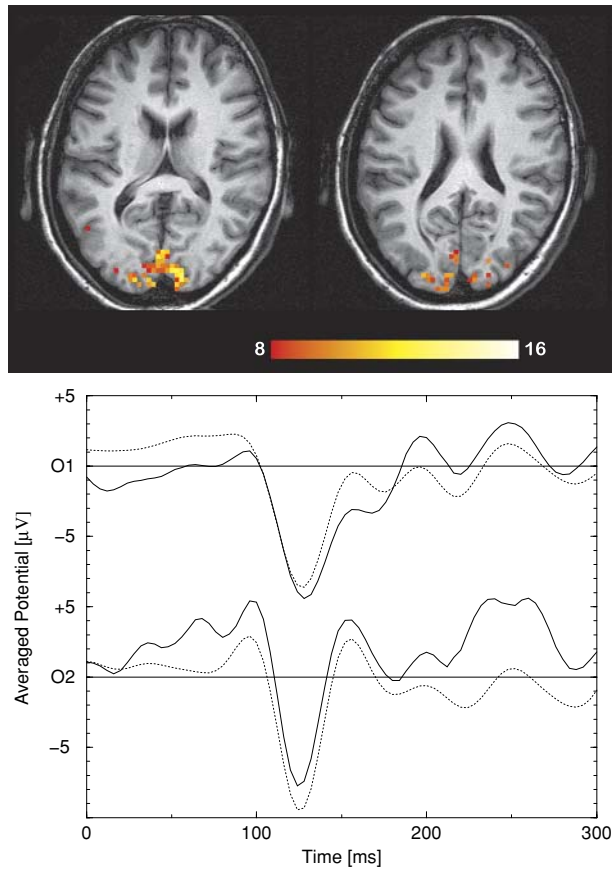


Figure 10.4 Example results from a single subject: BOLD activation (top) shown as a z-score color map overlaid onto the corresponding anatomical slices and visually evoked potential (below). Here, the dotted line corresponds to an experiment without fMRI scanning, the solid line results from a combined ERP/fMRI measurement.

study (5 female, 7 male, mean age 24.8 years, range 22-30 years). We used Kanizsa figures and non-Kanizsa figures (see Fig. 10.5) as stimulus material. Stimuli were presented for 1000 ms, followed by randomized inter-stimulus-intervals (ISI) of 1000 to 2500 ms. The ISI duration followed an exponential distribution corresponding to $ISI = 1000 - 500 * log(d), d \in [0.04979; 1]$. Figures were displayed in black on a white background with a black fixation cross in the center. Stimuli subtended a visual angle of 4.28 degrees including inducer disks.

The induced illusory figures (Fig. 10.5 left) subtended 2.86 degrees. Fixation crosses were displayed foveally (0.02 degrees). The ratio of the inducing line ends and the side-length of the illusory figures was 1/4. A block of 20 trials (approx. 50 s) was followed by a 10s display of the fixation cross alone. Fourty-five blocks were recorded in three experimental runs (a total of 900 trials). Conditions Kanizsa square (KS),

Figure 10.5 Stimulus material used in the experiment: on the left, Kanizsa square (KS) and triangle (KT), on the right: non-Kanizsa square (NS) and triangle (NT).

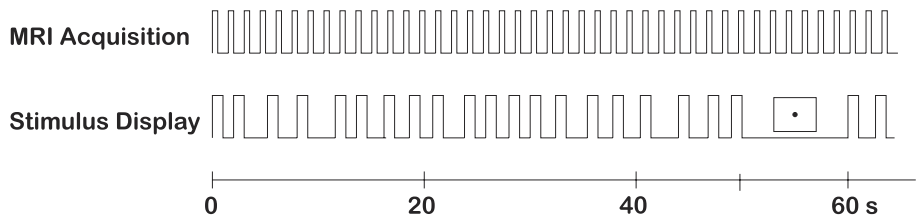


Figure 10.6 Scheme of experiment 2: Twenty Kanizsa stimuli were presented during 50 s with randomly jittered interstimulus intervals, followed by a 10 s display of a fixation cross. A clustered acquisition of 5 MRI slices during a period of 270 ms was performed at a repetition time of 1500 ms, leaving a 1230 ms period for EEG acquisition.

Kanizsa triangle (KT), non-Kanizsa square (NS), and non-Kanizsa triangle (NT) were presented equiprobably and randomized across subjects and runs. The Kanizsa square (KS) served as the target condition. Subjects were instructed to press a button with their right middle finger when a target appeared (p = 0.25), and to press another button with the right index finger for all other conditions (p = 0.75).

10.4.2.1 Data Recording

The EEG was recorded using the same set-up as above, and fMRI scanning was performed using the same equipment, but applying a slightly modified protocol (TE 30 ms, TR 1500 ms, 5 axial slices with thickness 6 mm, oriented parallel to the AC-PC line in the sagittal plane approx. at axial coordinates -13 mm, -5 mm, +3 mm, +43 mm, +50 mm. The time period during which the images were acquired was 270 ms, leaving a period of 1230 ms for EEG acquisition.

10.4.2.2 EEG Data Analysis

Collected EEG data were analyzed offline by the procedure described above. Finally, the corrected EEG was averaged across subjects in a period of -100 to +600 ms relative to stimulation onset, selecting periods with correct responses and specific conditions only (see Fig. 10.7).

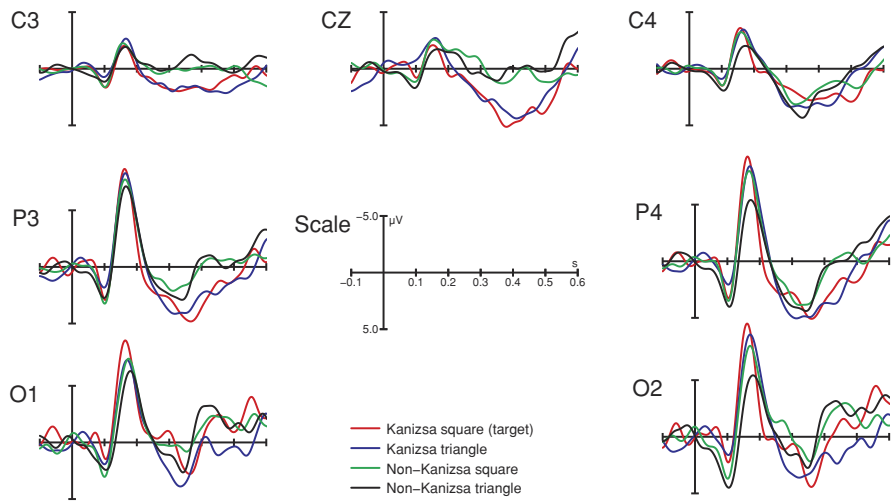


Figure 10.7 ERPs recorded at six selected positions for the four conditions Kanizsa square (KS, red), Kanizsa triangle (KT, blue), non-Kanizsa square (NS, green), non-Kanizsa triangle (NT, black).

As apparent from Fig. 10.7, all stimuli evoked the typical P100 and N170 ERP responses. For statistical analysis, ERP amplitudes were pooled into 6 regions: LA (left anterior: Fp1, F3), LC (left central: C3, T3), LP (left posterior: P3, O1), and their homologues on the right side. ERP components were defined by the time intervals 30-60 ms (N50), 70-110 ms (P100), 130-180 ms (N170), and 300-500 ms (P300). Repeated measures ANOVAs with factors topography (anterior, central, posterior), hemisphere (left, right), form (KS, KT, NS, NT) were conducted to assess the effect of the experimental variables on the measured amplitudes.

For the N50 component in both posterior regions, square figures elicited a higher amplitude than triangular ones ($F = 50.97, R^2 = 0.168, p < 1e-12$). No effect was found the the figure factor or the other regions. A similar finding was obtained for the P100 component ($F = 7.8, R^2 = 0.019, p = 0.05$). As might be inferred from Fig. 10.7, for the N170 component and both posterior regions, a significant ordering of amplitudes by the factor form was found: $KS > KT > NS > NT$ ($F = 31.7, R^2 = 0.182, p < 1e-12$). Likewise, the same ordering was found for the P300 component ($F = 13.3, R^2 = 0.022, p = 1e-8$). All results match with previous ERP (Herrmann et al., 1999) and MEG (Herrmann et al., 2000a) studies.

10.4.2.3 fMRI Data Analysis

After the same preprocessing steps as in the first experiment, fMRI data were analyzed by voxelwise univariate regression analysis. The regression analyses were designed to distinguish (1) task-related activation (KS, KT, NS, NT) from baseline (display of fixation cross alone), (2) target (KS) vs. non-target (KT, NS, NT)

related activation, (3) activation related to Kanizsa (KS, KT) vs. non-Kanizsa figures (NS, NT), and (4) activation related to squares (KS, NS) vs. triangles (KT, NT). In all designs, the first two time steps of each stimulus and baseline period were excluded from analysis as transition phases. In addition, the first 5 time steps of each scan were excluded due to their magnetical non-equilibrium. The design matrix was shifted by 5.5 s to match the lag of the hemodynamic response. Resulting z-score maps were registered with a T_1-weighted high resolution MR data set of the same subject and transformed into Talairach space, and averaged within the subject group. The resulting z-score map was thresholded by 4. Anatomical locations of activated areas are summarized in Tab. 10.1, and quantitative results from regression analyses are compiled in Tab. 10.2.

Table 10.1 Name and positions of activated regions-of-interest in experiment 2

Anatomical Location	ROI	Coordinates		
		x	y	z
Motor cortex left	MCL	-38	-19	53
Supplementary motor area	SMA	-5	1	54
Superior parietal lobule left	SPLL	-32	-53	52
Superior parietal lobule right	SPLR	29	-42	44
Middle frontal gyrus left	MFGL	-36	29	32
Middle frontal gyrus right	MFGR	33	45	24
Occipito-medial and lateral gyri left	OMGL	-42	-78	11
Occipito-medial and lateral gyri right	OMGR	39	-64	6
Precuneus (left and right)	PC	5	-33	51
Area striata (left and right)	AS	-10	-72	11
Heschl's gyrus left	HGL	-37	-17	- 4
Heschl's gyrus right	HGR	39	-7	- 0

Results of regression analyses were summarized as:

- During the stimulation (condition KS, KT, NS and NT vs. fixation point) activations are found at expected locations (see Tab. 10.1 and Tab. 10.2): the left motor cortex (MCL), the supplementary motor area (SMA), the left and right superior parietal lobule (SPLL, SPLR), bilateral occipito-medial (OML, OMR) and occipito-lateral gyri (OLL, OLR). Interestingly, the periphery of the area striata (AS) and Heschl's gyrus is suppressed on both sides. This is interpreted as an attentional focusation on the center of the visual field while suppressing the peripheral visual field and the primary auditory cortex.
- The evaluation of target (KS) vs. non-target (KT, NS, NT) conditions revealed an activation of a bilateral fronto-parietal network (MFGL, SPLL, MFGR, SPLR),

Table 10.2 Activation foci

ROI	(KS, KT, NS, NT) vs. Fixation		(KS) vs. (KT, NS, NT)		(KS, KT) vs. (NS, NT)		(KS, NS) vs. (KT, NT)	
	Integral	z_{max}	Integral	z_{max}	Integral	z_{max}	Integral	z_{max}
MCL	40181	11.81	26305	8.97	21618	9.10	14428	8.35
SMA	5787	10.16	6788	6.24	3586	6.74	1856	6.39
SPLL	2654	8.61	1453	5.31	–	n.s.	–	n.s.
SPLR	1354	7.56	79	4.36	–	n.s.	–	n.s.
MFGL	-520	-5.93	561	5.50	–	n.s.	–	n.s.
MFGR	-10122	-8.80	200	4.68	–	n.s.	–	n.s.
OMGL	52288	14.58	–	n.s.	12054	7.82	6448	7.22
OMGR	42609	14.84	–	n.s.	13520	9.11	1956	6.95
PC	-20091	-11.62	–	n.s.	–	n.s.	–	n.s.
AS	-244861	-19.24	30305	6.82	-1065	-6.41	1967	7.24
HGL	-3068	-7.49	–	n.s.	–	n.s.	–	n.s.
HGR	-19146	-9.54	–	n.s.	–	n.s.	–	n.s.

Notes: Selected activation foci for the all stimuli vs. fixation, target vs. non-target conditions, Kanizsa vs. non-Kanizsa figures and square vs. triangle figures. For each comparison, the integral suprathreshold activation and the maximum $z-$score within a focus are given.

and a stronger activation of MCL, SMA and SPL. The periphery of AS exhibits a relative activation (i.e., a less pronounced suppression).

- Kanizsa figures elicit a stronger activation of MCL and SMA, however less pronounced compared to the target condition. The detection of "meaningful figures" is documented by relatively stronger activations of secondary visual areas.
- Displaying squares (KS, NS) elicits a stronger activation of MCL, SMA, OMG and OLG than triangles (KT, NT). However, this effect is less pronounced compared to the target effect and the effect elicited by the Kanizsa figures. The periphery of AS is less suppressed, which might be explained by the larger spatial extent of the squares.
- In summary, the activation increases in MCL and SMA for squares, Kanizsa figures, and the target condition. OMG and OLG exhibit a stronger activation for squares and Kanizsa figures. SPL appears to be involved in the detection of the target condition. The periphery of AS is suppressed during stimulus display, which is less pronounced for squares and target display.

10.4.2.4 Modelling the Hemodynamic Response

To obtain information about the class-wise shape properties of the hemodynamic response, a non-linear regression model was adapted to the time-series (Kruggel

et al., 2000b). Each hemodynamic response due to a single stimulus is a modeled by a Gaussian function. We assume that each stimulus of a given class elicits the same response, and that subsequent stimuli add linearly (Buckner et al., 1996). This defines the model of the time-series y as:

$$y = \sum_{s} \sum_{t=0}^{t_{max}} (g_{c(s)} * \exp(-((t - l_{c(s)})/d)^2) + o, \tag{10.1}$$

where the parameters of the Gaussian function are called g: gain, l: lag, d: dispersion and o: offset. The inner sum models the hemodynamic response due to a single stimulus in the time interval $t \in [0, t_{max}]$ lasting from stimulation onset for an arbitrary time (here, $t_{max} = 12s$). The outer sum runs over all trials s of the experiment, with $c(s) \in \{KS, KT, NT, NS\}$ referring to the stimulus class. Note that the dispersion and the offset were assumed as class-independent. The signal was sampled at integral multiples of TR (the actual time points of measurements), and the slice-dependent acquisition delay of the EPI protocol was taken into account.

First, regions-of-interest (ROIs) were determined by computing a regression analysis in single subjects as described above, measuring the effect of stimulation periods (KS, KT, NS, NT) vs. fixation point display. In the resulting individual $z-$score maps, we defined regions of 6 four-connected, suprathreshold ($z \geq 6$) voxels around local maxima and selected those regions, whose position most closely resembled to regions found in the group analysis (see Tab. 10.2). The time series for a ROI was obtained by averaging voxel intensities at a given timepoint. Note that spatio-temporal correlations were neglected.

Parameters of the model function were optimized using Powell's algorithm (Press et al., 1992). Ten parameters (gain and lag for each class, class independent dispersion and offset) were determined from a time series of 1800 points. For inter-subject comparisons, relative gain values were computed for each subject and each ROI: $rg_c = g_c / \sum_c g_c$. Lag times were normalized by subtracting the individual lag of a ROI within the area striata (AS).

For each subject, ROI and stimulus class, we obtained a relative gain (activation strength) and a relative lag (time to response maximum). Resulting values were ordered by time and condition. Orderings were determined by computing Student's t tests (single sided, unequal variance, where $>$ corresponds to $p < 0.05$, $\sim$ to $p \geq 0.05$):

- The temporal sequence of activations of ROIs and their mean relative lag was determined as: MCL (-0.019 s) $\sim$ AS (0.000 s) $\sim$ OMG (0.034 s) $\sim$ OLG (0.106 s) $<$ SMA (0.515 s) $<$ SPL (+1.100 s). Three temporal activation groups are discriminated: (AS, OMG, OLG, MCL) appear first, then SMA, then SPL.
- Lag times for the target condition tended to be greater for the target condition in ROIs SPL (Δt = 0.380 ms, p = 0.079) and SMA (p = 0.056), but not in the other ROIs.
- Relative gains vs. experimental conditions were ordered for all ROIs. We obtained for ROI AS: KS $\sim$ NS $\sim$ KT $\sim$ NS, for ROIs OMG, OLG: KS $\sim$ KT $>$ NT $\sim$ NS,

for ROIs MCL, SMA and SPL: KS > NT ∼ KT ∼ NS. While the activation of central portions of the striate cortex was independent from the stimulus, a stronger activation was found for Kanizsa figures in OMG and OML. A clear selection of the target was found in ROIs MCL, SMA and SPL.

Activations in ROIs MFGL and MFGR were too low to warrant a proper modelling.

10.4.2.5 Combining Results

We would now try to summarize results from this combined ERP-fMRI study of an oddball task using illusory figures:

- ERP responses in the N50 and P100 time window from this and an earlier experiment (Herrmann et al., 1999) demonstrated a slightly higher activation for squares than for triangles. This result is consistent with a slightly higher activation of the striate cortex found by the standard linear and the non-linear regression model. We argue that this due to that fact that the four pac-men, when flashed over the screen, lead to a greater change in overall brightness and have a greater spatial extent than only three of them.
- At a later processing stage (during the N170 time window), Gestalt-like properties that emerge from binding individual elements seem to become more relevant (Herrmann and Bosch, 2001). The target effect in striate cortex is probably not due to an early selection mechanism since the early ERPs (N50, P100) do not show a target effect (Heinze et al., 1994). It is more likely that the striate cortex receives feedback from higher visual areas during a later stage of the selection (Martinez et al., 1999). The ERP component indicated an amplitude-ordering by condition as: KS > KT > NS > NT, which is consistent with the regression results for ROIs OML and OLL.
- A fronto-parietal network (consisting of ROIs SPL and MFG bilaterally) is responsible for response selection: Both regions are more strongly activated when comparing target vs. non-target conditions in fMRI; this and previous EEG and MEG studies (Herrmann et al., 1999, 2000a) have demonstrated maxima of the P300 component over the centro-parietal cortex (at positions P3, P4, Herrmann et al. (1999). The locations of the BOLD effect which resemble this effect (OMG, OLG) are in accordance with previous fMRI localization of the N170 sources (Gonzales et al., 1994).
- Similar to results from modelling other fMRI experiments (Kruggel and von Cramon, 1999; Kruggel et al., 2000b), ROIs SMA and MCL, which are responsible for response generation, are activated rather early and show a clearly stronger activation during target trials: KS > NT ∼ KT ∼ NS.
- The auditory cortex and periphery of the striate cortex are suppressed during stimulation. Most likely, this corresponds to an attentional focusation on central visual cortex.

This qualitative hypothetical model is in accordance with experimental findings from this combined and previous separate measurements. It implicitly reflects the underlying hypothesis that the absolute ERP amplitudes correspond to fMRI activation strengths, although there is no strict experimental evidence for this assumption.

10.5 Discussion

The feasibility of recording ERPs during fMRI scanning using cognitive stimuli was demonstrated by recording ERPs with the expected configuration while measuring a typical pattern of BOLD responses. While the possibilities of this new methodology are exciting, a few issues should be remembered when planning such experiments, analyzing their data, or interpreting their results:

Problems of a combined measurement: There are mutual influences of the EEG and MRI measuring process. The clustered EPI protocol in experiment 2 allowed recording 5 functional slices in 250 ms, but the EEG amplifier needed approx. 150 ms to recover from saturation. Thus, a window of 400 ms is lost for each block of scans from the EEG time course. For an average trial duration of 2.5 s here, this compiles to an acceptable "duty cycle" of 84 %. However, if the process under study requires scanning of a larger extent of the brain, this might leave EEG windows left which are too short for a meaningful evaluation.

The cardio-ballistic effect may be corrected by using one of the published methods (Allen et al., 1998, 2000; Kruggel et al., 2000a). Due to their comparatively high magnitude on our 3T scanner, remnants of this artifact are still detectable in the corrected output, which corresponds to a lower SNR ratio in the grand averages. We estimate that 2-3 times the trials of a conventional ERP experiments are needed in a combined EEG-fMRI measurement. To avoid picking up too much of this pulse-synchronous signal, we had to move the reference from the Goldmann point to the nasion. However, this drastically reduced the relative amplitude of the P300 component for the target condition.

Using 20 EEG electrodes and cables, which were radially joined at location Pz of the electrode cap, resulted in a loss of MR signal which was most noticable in the topmost slices. This is best explained by a shielding effect of the cables, and made the shimming process of the MR scanner tedious. Using this conventional EEG set-up, this certainly poses an upper limit for the number of electrodes, most likely not much beyond 20.

Problems of exprimental design: As stated above, the rather low SNR of the grand average forces the design of experiments with a rather high number of trials per class (say, at least 100). On the other hand, most detail about the shape properties of the hemodynamic response is obtained when using rather long trial lengths (say, 12 s or more), so that the overlap of sequential BOLD responses is negligible. Obviously, a compromise between the number of stimulus classes in a factorial (or parametric) design, and the trial length must be made.

We employed the rapid stimulation protocol using randomly varied trial length introduced by Burock et al. (1998) in the second experiment. We presented 900 trials within 45 min, using a trial length of 2–3.5 s, and we were able to disentangle the class-wise properties of hemodynamic response by non-linear regression analysis. Obviously, such rapid presentation is better suited to visual than auditory stimulus material.

A greater freedom in experimental design is possible when improving the SNR of ERP acquistion to a level comparable to separate measurements. An improved subject set-up, and advances in sensor and amplifier construction are expected to yield most benefit. Because artifacts scale supralinearly with field strengths, it might even be beneficial to conduct such experiments at 1.5 Tesla field strengths, sacrificing sensitivity for the BOLD signal in favor of a less-distorted EEG.

Problems of analyzing data: We analyzed measured EEG and fMRI data in a conventional fashion, i.e., each measurement separately. When trying to create a synthesis of the results for ERP and fMRI data analysis, the following physiological response properties must be remembered: Electric potentials measured on the scalp have a rather low spatial resolution, which is partially due to a spatial low-pass filtering effect of the outer hulls of the brain, and thus correspond to an integral response of a certain brain region at a given time point. Conversely, the BOLD response may be understood as a (fast) neuronal activation convolved by a (slow) hemodynamic response function. This corresponds to a low-pass filtering effect in time, or: an integral activation over a certain time window at a specific brain location. In addition, it is still unclear to which extent lag times are influenced by a delay in neuronal activation *or* in delivery of oxygenated blood to the response area.

Theoretical considerations: Although both experiments demonstrate the feasibility of conducting combined experiments, theoretical considerations warn against a naive interpretation of the results. As Nunez and Silberstein (2000) pointed out, it is important to remember that:

- EEG and fMRI responses may not necessarily originate from the same cell assemblies, and
- Scalp EEG amplitudes and hemodynamic activity can change in opposite direction.

Much of the physiological knowledge necessary to create a computational model of the neurono-vascular coupling is still missing, which is required to permit a stronger interpretation of the results revealed by combined experiments. The cooperativity of the brain adds another level of complexity when interpreting results: Assuming that processes are strictly sequential in time and well-separated in space, unique models about the underlying processes involved in a cognitive task may be constructed. While this assumption may approximately hold for early processing stages (e.g., during stimulus perception), it is well known that later processing stages (i.e., stimulus analysis, decision making, response generation) require a network of temporally strongly overlapping processes (Makeig et al., 1999), where

even re-activations of certain brain regions are under discussion (Martinez et al., 1999). Modelling such a network will most likely yield non-unique solutions.

We demonstrated the feasibility of performing combined ERP/fMRI experiments under higher-level cognitive stimulation. The perspective of observing complementary responses due to the same stimulation event on a single subject level is very appealing in order to better understand physiological processes underlying brain activation and the functional organization of the brain.

References

P.J. Allen, O. Josephs, and R. Turner. A method for removing imaging artifact from continuos EEG recorded during functional MRI. *NeuroImage*, 12:230–239, 2000.

P.J. Allen, G. Polizzi, K. Krakow, D.R. Fish, and L. Lemieux. Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction. *NeuroImage*, 8:229–239, 1998.

A.M. Dale, B.R. Rosen, and R.L. Buckner. Event-related functional MRI: Past, present and future. *Proceedings of the National Academy of Sciences USA*, 95: 773–780, 1998.

R.L. Buckner. Event-Related fMRI and the hemodynamic response. *Human Brain Mapping*, 6:373–377, 1998.

R.L. Buckner, P.A. Bandettini, K.M. O'Craven, R. Savoy, S.E. Petersen, M. Raichle, and B. Rosen. Detection of transient and distributed cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences USA*, 93:14878–14883, 1996.

M.A. Burock, R.L. Buckner, M.G. Woldorff, B.R. Rosen, and A.M. Dale. Randomized event-related experimental designs allow for extremly rapid presentation rates using functional MRI. *Neuroreport*, 9:3735–3739, 1998.

C.G. Celesia and M.G. Brigell. Recommended standards for pattern electroretinograms and visual evoked potentials. *Electroencephalography and Clinincal Neurophysiology*, 52 (Suppl):53–67, 1999.

V.P. Clark, J.M. Maisog, and J.V. Haxby. FMRI study of face perception and memory using random stimulus sequences. *Journal of Neurophysiology*, 79:3257–3265, 1998.

K.J. Friston, S. Williams, R. Howard, R.S.J. Frackowiak, and R. Turner. Movement-related effects in fMRI time series. *Magnetic Resonance in Medicine*, 35:346–355, 1996.

K.J. Friston, K.J. Worsley, R.S.J. Frackowiak, J.C. Mazziotta, and A.C. Evans. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, 1:210–220, 1994.

R.I. Goldman, J.M. Stern, J. Engel, and M.S. Cohen. Acquiring simultaneous EEG and functional MRI. *Clinical Neurophysiology*, 111:1974–1980, 2000.

C.M. Gomez Gonzales, V.P. Clark, S. Fan, S.J. Luck, and S.A. Hillyard. Sources of attention-sensitive visual event-related potentials. *Brain Topography*, 7:41–51, 1994.

H.J. Heinze, G.R. Mangun, W. Bruchert, H. Hinrichs, M. Scholz, T.F. Münte, A. Gös, S. Johannes, M. Scherg, H. Hundeshagen, M.S. Gazzaniga, and S.A. Hillyard. Combined spatial and temporal imaging of spatial selective attention in humans. *Nature*, 392:543–546, 1994.

C.S. Herrmann and V. Bosch. Gestalt perception modulates early visual processing. *Neuroreport*, 12:901–904, 2001.

C.S. Herrmann, A. Mecklinger, and E. Pfeifer. Gamma responses and ERPs in a visual classification task. *Clinical Neurophysiology*, 110:636–642, 1999.

C.S. Herrmann, A. Mecklinger, and E. Pfeifer. Magnetoencephalographic responses to illusory figures: Early evoked gamma is affected by processing of stimulus features. *International Journal of Psychophysiology*, 38:265–281, 2000a.

C.S. Herrmann, U. Oertel, Y. Wang, B. Maess, and A.D. Frederici. Noise affects auditory and linguistic processing differently: an MEG study. *Neuroreport*, 11: 227–229, 2000b.

R.J. Ives, S. Warach, F. Schmitt, R.R. Edelmann, and D.L. Schmoer. Monitoring the patient's EEG during echo planar MRI. *Electroencephography and Clinical Neurophysiology*, 87:417–420, 1993.

K. Krakow, P.J. Allen, M.R. Symms, L. Lemieux, O. Josephs, and D.R. Fish. EEG recording during fMRI experiments: Image quality. *Human Brain Mapping*, 10: 10–15, 2000.

F. Kruggel, C.S. Herrmann, C.J. Wiggins, and D.Y. von Cramon. Hemodynamic and Electroencephalographic responses to illusory figures: Recording of the Evoked Potentials During Functional MRI. *Neuroimage*, 14:1327–1336, 2001.

F. Kruggel and D.Y. von Cramon. Modeling the hemodynamic response in single-trial functional MRI experiments. *Magnetic Resonance in Medicine*, 42:787–797, 1999.

F. Kruggel, D.Y. von Cramon, and X. Descombes. Comparison of filtering methods for fMRI datasets. *NeuroImage*, 10:530–543, 1999.

F. Kruggel, C.J. Wiggins, C.S. Herrmann, and D.Y. von Cramon. Recording of the event-related potentials during functional MRI at 3.0 Tesla field strength. *Magnetic Resonance in Medicine*, 44:277–282, 2000a.

F. Kruggel, S. Zysset, and D.Y. von Cramon. Nonlinear regression of functional MRI data: An item recognition task study. *NeuroImage*, 12:173–183, 2000b.

L. Lemieux, P.J. Allen, F. Franconi, M.R. Symms, and D.R. Fish. Recording of EEG during fMRI experiments: Patient safety. *Magnetic Resonance in Medicine*, 38:943–952, 1997.

A.K. Liu, J.W. Belliveau, and A.M. Dale. Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulation. *Proceedings of the National Academy of Sciences USA*, 95: 8945–8950, 1998.

S. Makeig, M. Westerfield, J. Townsend, T.P. Jung, E. Courchesne, and T.J. Sejnowski. Functionally independent components of early event-related potentials in a visual spatial attention task. *Philosophical Transactions of the Royal Society: Biological Sciences*, 354:1135–1144, 1999.

A. Martinez, L. Anllo-Vento, M.I. Sereno, L.R. Frank, R.B. Buxton, D. Dubowitz, E.C. Wong, H. Hinrichs, H.J. Heinze, and S.A. Hillyard. Involvement of striate and visual cortical areas in spatial attention. *Nature Neuroscience*, 2:364–369, 1999.

R.M. Müri, J. Felbinger, K.M. Rösler, B. Jung, C.W. Hess, and C. Boesch. Recording of electrical brain activity in a magnetic resonance environment: Distorting effects of the static magnetic field. *Magnetic Resonance in Medicine*, 39:18–22, 1998.

P.L. Nunez and R.B. Silberstein. On the relationship of synaptic activity to macroscopic measurements: Does co-registration of EEG with fMRI make sense? *Brain Topography*, 13:79–96, 2000.

B. Opitz, A. Mecklinger, D.Y. von Cramon, and F. Kruggel. Combining electrophysiological and hemodynamic measures of the auditory oddball. *Psychophysiology*, 36:142–147, 1999.

W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge and MA, 1992.

W. Richter, P. Andersen, A.P. Georgopoulos, and S.G. Kim. Sequential activity in human motor areas during a delayed cued finger movement task studied by time-resolved fMRI. *Neuroreport*, 8:1257–1261, 1998.

M.G. Woldorff, C. Tempelmann, J. Fell, C. Tegeler, B. Gaschler-Markefski, H. Hermann, H.J. Heinze, and H. Scheich. Lateralized auditory spatial perception and the contralaterality of cortical processing processing as studied with fMRI and MEG. *Human Brain Mapping*, 7:49–66, 1999.

K.J. Worsley and K.J. Friston. Analysis of fMRI time-series revisited-again. *NeuroImage*, 2:173–181, 1996.

This page intentionally left blank

IV Models Integrating Neurophysiology and Functional Imaging

This page intentionally left blank

11 Synthetic PET Imaging for Grasping: From Primate Neurophysiology to Human Behavior

Michael A. Arbib
USC Brain Project
University of Southern California
Los Angeles, California 90089-2520
arbib@pollux.usc.edu

Andrew H. Fagg
Computer Science Department
University of Massachusetts
Amherst, MA

Scott T. Grafton
Dartmouth College
Hanover, NH

Synthetic PET imaging is a technique for using computational models derived from primate neurophysiological data to predict and analyze the results of human PET studies. This technique makes use of the hypothesis that is correlated with the integrated synaptic activity in a localized brain region. In this chapter, we describe the Synthetic PET imaging approach, and demonstrate how it is applied to the FARS model of parietal-premotor interactions underlying primate grasp control. The Synthetic PET measures are computed for a simulated conditional/non-conditional grasping experiment, and then compared to the results of a similar human PET study. We then show how the human PET results may be used to further constrain the computational model.

11.1 Synthetic PET Defined

In order to provide a causal account of brain function constrained by data from both primate neurophysiology and human brain imaging, Arbib et al. (1995) introduced a new computational technique, called Synthetic PET imaging. This technique uses neural models that are based on primate neurophysiology to predict and analyze results from PET (Positron Emission Tomography) brain imaging regional cerebral blood flow (rCBF) or glucose metabolism taken during performance of a variety of human behaviors. The problem is to find an integrated measure of activity in each simulated neural group that provides a predictor for the PET-measured activation of the 3D volume to which the neurons in this group correspond. The key hypothesis is that PET metabolic imaging is correlated with the integrated synaptic activity in a brain region (Brownell et al., 1982), and thus reflects in part neural activity in regions afferent to the region studied, rather than intrinsic neural activity of the region alone. However, the method is general, and can potentially accommodate other hypotheses on single cell correlates of imaged activity, and can thus be applied to other imaging techniques, such as functional MRI, as they emerge (see Arbib et al. (1995) for further discussion). Thus, although the present study uses Synthetic PET, we emphasize that this is but one case of the broader potential for systems neuroscience of Synthetic Brain Imaging (SBI) in general. In the rest of this section we briefly review the way in which we represent neural networks for computer simulation, and then provide the formal definition for Synthetic PET.

11.1.1 Modeling Neural Networks

Here, we adopt the leaky integrator neuron model leaky integrator model of the neuron, in which the internal state of the neuron is described by a single variable, the membrane potential m(t) at the spike initiation zone. The time evolution of m(t) is given by the differential equation:

$$\tau \frac{dm(t)}{dt} = -m(t) + \sum_i w_i X_i(t) + h, \tag{11.1}$$

with resting level h, time constant τ, $X_i(t)$ the firing rate at the ith input, and w_i the corresponding synaptic weight. The present model defines the firing rate as a continuously varying measure of the cell's activity. The firing rate is approximated by a sigmoid function of the membrane potential, $M(t) = \sigma(m(t))$[1]. Many brain regions can be modeled as a set of two dimensional arrays of neurons, with one array for each anatomically or physiologically distinct cell type. Connections between these neural arrays are defined in terms of interconnection masks which describe

1. An appreciation of neural complexity is necessary for the computational neuroscientist wishing to determine how detailed the neural model needs to be when studying a specific system—see Rall (1995) and Arbib (1995), for further details.

the synaptic weights. E.g., the equations $\tau = 10$ ms, and $S_A = C + W * B$, state that the membrane time constant for neural region A, τ is 10 milliseconds, and that for each cell i,j in array A, the cell's input, $S_A(i,j)$, is the sum of the output of the i,j^{th} cell in C, plus the sum of the outputs of the 9 cells in B centered at i,j times their corresponding weights in W. In other words,

$$S_A(i,j) = C(i,j) + \sum_{k,l=-l}^{l} W(k,l) \cdot B(i+k, j+1),$$

that is, the $*$ operator in "$W * B$" indicates that mask W is spatially convolved with B.

11.1.2 Defining Synthetic PET

The issue now is how to map the activity simulated in neural network models of interacting brain regions based on say single-cell recordings in behaving monkeys into predictions of metabolic activity values to be recorded from corresponding regions of the human brain by imaging techniques such as PET. There are two problems: localization and modeling activation.

Localization: Each array in the neural network model represents a neural population in a region identified anatomically and physiologically in the monkey brain. A Synthetic PET comparison requires explicit hypotheses stating that each such region A is homologous to a region h(A) in the human brain such that —within the tasks under consideration— A and h(A) perform their tasks in the same way. In some cases, such homologies are well defined. In other cases, the existence or identity of such a homology is an open question. Thus, the comparison of a Synthetic PET study with the results of a human brain scan study will, inter alia, be a test of the hypothesis "h(A) in human is homologous to A in (a given species of) monkey", and comparison of synthetic and human studies may suggest a new homology to be tested in further studies.

Modeling activation: PET typically measures regional cerebral blood flow (rCBF). Arbib, et al. (1995) hypothesize that the counts acquired in PET scans are correlated with local synaptic activity in a particular region (Brownell et al., 1982; Fox and Raichle, 1985), and call this measure the "raw PET activity." However, PET studies typically do not report these values, but instead report the comparative values of this activity in a given region for two different tasks or behaviors.

We thus define our Synthetic PET computation in two stages:

1. Compute $rPET_A$, the simulated value of raw PET activity, for each region A of our network while it is used to simulate the monkey's neural activity in some given task.

2. Compare the activities computed for two different tasks. The result is a Synthetic PET comparison which presents our prediction of human brain activity as based on neural network modeling constrained by monkey neurophysiology and known functional neuroanatomy.

The synthetic raw activity, $rPET_A$, associated with a cell group A is defined as:

$$rPET_A = \int_{t_0}^{t_1} \sum_B \omega_{B\to A}(t)dt, \tag{11.2}$$

where A is the region of interest, the sum is over all regions B that project to the region of interest, $w_{B\to A}(t)$ is the synaptic activity ($firing\ rate * |synaptic\ strength|$) summed over all the synapses from region B to region A at time t, and the time interval from t_0 to t_1 corresponds to the duration of the scan (see Arbib et al. (1995) for further discussion).

The comparative activity $PET_A(1/2)$ for task 1 over task 2 for each region A was given by Arbib et al. (1995) as:

$$PET_A(1/2) = \frac{rPET_A(1) - rPET_A(2)}{rPET_A(1)} \tag{11.3}$$

where $rPET_A(i)$ is the value of $rPET_A$ in condition i, to compare the change in PET_A from task 2 to task 1. In the present study we use a different measure, defining the *relative synaptic activity* for region A from task 1 to task 2 with $max(rPET_A(1), rPET_A(2))$ replacing $rPET_A(1)$ in the denominator of equation 11.3 to yield

$$PET_A(1/2) = \frac{rPET_A(1) - rPET_A(2)}{\max_{i\in\{1,2\}} rPET_A(i)} \tag{11.4}$$

This yields a more robust measure of relative synaptic activity. We can display the values of the "Synthetic PET comparison" $PET_A(1/2)$ for each region A on a graph or in a table, or we may (see Arbib et al. (1995)) convert each A-value to a color scale, and display the colors on the region $h(A)$ homologous to A on slices based on the Talairach Atlas (Talairach and Tournoux, 1988). The resulting images then predict the results of human PET studies. Note that we are plotting *synaptic* activity for each region A, not the neural activity of A. As a computational plus (going beyond the imaging technology), we may also collect the contributions of the excitatory and inhibitory synapses separately, based on evaluating the integral in equation 11.2 over one set of synapses or the other. Using simulated PET, we can break apart different factors that contribute to the measure of synaptic activity so that they can be studied independently. This can allow a much more informed view of the actual PET data that are collected, possibly shedding light on apparent contradictions that arise from interpreting rCBF simply as cell activity.

11.2 A Model of Grasp Control

The cells of area F5, part of of the macaque inferior premotor cortex, are often selective for the type of grasp made by the monkey (Rizzolatti et al., 1988). Grasps observed during these experiments include precision pinches (using the tips of the index finger and thumb), lateral pinches (thumb against the side of the index

finger), and power grasps (four fingers opposing the palm). In addition, the firing of these cells typically correlated with a particular phase of the ongoing movement. For a task in which the monkey was presented with an object, then grasped the object in response to a go signal, held the object, and finally released the object after a secondary go signal, the following phases were identified: preparatory (set), finger extension, finger flexion, holding, and release. F5 exchanges cortico-cortical connections with AIP (the anterior intra-parietal area of parietal cortex), whose cells demonstrate a variety of both visual- and grasp-related responses (Taira et al., 1990). This section outlines the FARS (Fagg-Arbib-Rizzolatti-Sakata) model of the grasping process. It is implemented in terms of simplified, but biologically plausible neural networks. For details of the model, supporting monkey data, computational constraints, and a set of simulation results, see Fagg (1996) andFagg and Arbib (1998). In the next section, we will extract measures of regional synaptic activity from the model, and then compare them to rCBF results in a human PET study. The FARS model focuses on the roles of several intra-parietal areas (anterior - AIP, posterior - PIP, and ventral - VIP), inferior premotor cortex (F4 and F5), pre-SMA (F6; one of two subdivisions of the supplementary motor area), frontal cortex (area 46), F2 (dorsal premotor cortex), inferotemporal cortex (IT), the secondary somatosensory cortex (SII), and the basal ganglia (BG). However, in this chapter we shall discuss only the contributions of AIP, F5, F6, F2, and the BG. The crucial aspects of the model (see figure 11.1) are the following:

1. AIP serves the dual role of first computing a set of affordances for the object being attended (i.e., AIP highlights properties of the object relevant for manually interacting with it), and then maintaining an active memory of the selected affordance as the corresponding grasp is prepared and executed.

2. F5 integrates a variety of constraints to decide on the single grasp that is to be executed. These constraints include visual information (from the affordances extracted by AIP), task information (from F6), instruction stimuli (from F2), and a working memory (from area 46) of recently-executed grasps. We shall say more of F6 and F2 below; area 46 will not be considered further in the present paper. When the movement is triggered, F5 is responsible for the high-level execution and subsequent monitoring of the planned preshape and grasp.

Fagg and Arbib (1998) have offered both a computational analysis and an analysis of empirical data in support of the hypothesis that not only is F5 responsible for unfolding the grasp in time during the execution of the movement, but that F5 also sends recurrent connections back to AIP to update AIP's active memory for the grasp that is about to be executed or that is being executed by F5.

Figure 11.1 illustrates how the modeled AIP computes a set of affordances for a mug, and passes the corresponding set of grasps to F5. In general, a single object affords many possible grasps. As a function of the current context, F5 selects only one. This decision is then broadcast back to AIP, which shunts the other affordances, leaving only the affordance that corresponds to the selected grasp. During the execution of the grasp, the affordance represented by AIP forms the active memory

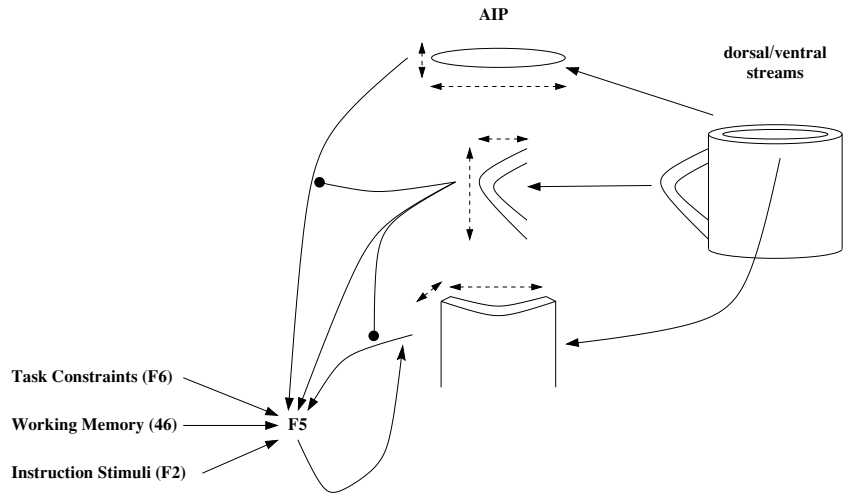


Figure 11.1 According to the FARS model, AIP uses visual input to extract affordances, which highlight the features of the object that are relevant to grasping it. F5 then applies various constraints to select a grasp for execution and to inform AIP of the status of its execution, thus updating AIP's active memory. The areas shown are AIP (anterior intraparietal cortex), area F5 (of the ventral premotor cortex), and regions providing supporting input to F5, namely F6 (pre-SMA), area 46 (dorsolateral prefrontal cortex), and F2 (dorsal premotor cortex).

which is continually updated by inputs from the active grasp program in F5. This process of separating out motor-related visual features may explain why cells in AIP reflect both object- and grasp-related activity patterns.

The *current context* used by F5 to select amongst available grasps may include task requirements, position of the object in space, and even obstacles. When the precise task is known ahead of time, it is assumed that a higher level planning region predisposes the selection of the correct grasp. In the FARS model, it is area F6 that performs this function. However, here we emphasize a task in which the grasp is not known prior to presentation of the object, and is only determined by an arbitrary instruction stimulus made available during the course of the trial (e.g. an LED whose color indicates one of two grasps). The dorsal premotor cortex (F2) is thought to be responsible for the association of arbitrary IS with the preparation of motor programs (Evarts et al., 1984; Kurata and Wise, 1988; Wise and Mauritz, 1985). In a task in which a monkey must respond to the display of a pattern with a particular movement of a joystick, some neurons in F2 respond to the sensory-specific qualities of the input, but others specifically encode which task is to be executed on the basis of the instruction, they thus form set cells which encode the motor specification until the go signal is received (Fagg and Arbib, 1992; Mitz et al., 1991).

We therefore implicate F2 as a key player in this grasp association task. What is particularly interesting about this type of conditional task, is that alone, neither the view of the object (with its multiple affordances), nor the instruction stimulus (IS)

is enough to specify the grasp in its entirety: the visual input specifies the details of all the possible grasps; the IS specifies only the grasp mode—and not the specific parameters of the grasp (such as the aperture). F5 must combine these sources of information in order to determine the unique grasp that will be executed.

11.2.1 Population Coding of Grasp Type in AIP and F5

Figure 11.2 presents an outline of the neural regions involved in the FARS model of grasp production. The precision pinch and power grasp pools of AIP receive inputs from both the dorsal and ventral visual pathways (pathways not shown; more details may be found in (Fagg , 1996; Fagg and Arbib, 1998). The pools in F5 and AIP are connected through recurrent excitatory connections: affordances represented by populations of units in AIP excite corresponding grasp cells in F5; active F5 units, representing a selected grasp, in turn support the AIP units that extract the motorically-relevant features of the objects. In monkey, the number of neurons in F5 involved in the execution of the precision pinch is greater than the number observed for any other grasp (Rizzolatti et al., 1988). The model reflects this distribution in the sizes of the precision and power pools in both F5 and AIP.

Each grasp pool within F5 is partitioned into overlapping subpopulations that encode both the phase of the grasp program, and the grasp parameters (e.g. the grasp aperture). Subpopulations within AIP capture the affordance parameters, such as object width. Cells within these subpopulations exchange excitatory connections with one another, supporting their mutual coactivation. Competition between opposing subpopulations is mediated via inhibitory interneurons (indicated by the connections in figure 11.2 that are terminated with filled circles). F5 cells that are active for a given phase of movement recruit units in the primary motor cortex (F1) that move the fingers in a manner that is appropriate for that phase. In addition, units in the secondary somatosensory cortex (SII) are recruited by F5 cells as a way of monitoring the progress of the grasp as it is executed. The results of this monitoring are broadcast back to F5, which may in turn adjust the ongoing execution of the program. When the model is presented with the conditional task described in the previous section, how is a unique grasp selected for execution? AIP first extracts the set of affordances that are relevant for the presented object (say, a cylinder). These affordances, which also encode the diameter of the cylinder, activate the corresponding motor set cells in F5. However, because there are multiple active affordances, several competing subpopulations of F5 set cells achieve a moderate level of activation. This competition is resolved only when the IS is presented. This instruction signal, mapped to a grasp mode preference by the basal ganglia (connections not shown in the figure 11.2), is hypothesized to arrive at F5 via F2. The signal increases the activation level of those F5 cells that correspond to the selected grasp, allowing them to win the competition over the other subpopulations. Besides the processing of instruction stimuli, the basal ganglia play two additional roles in the model. A subset of BG units are dedicated to implementing the gating circuitry that controls the phasic behavior of cells within F5. This

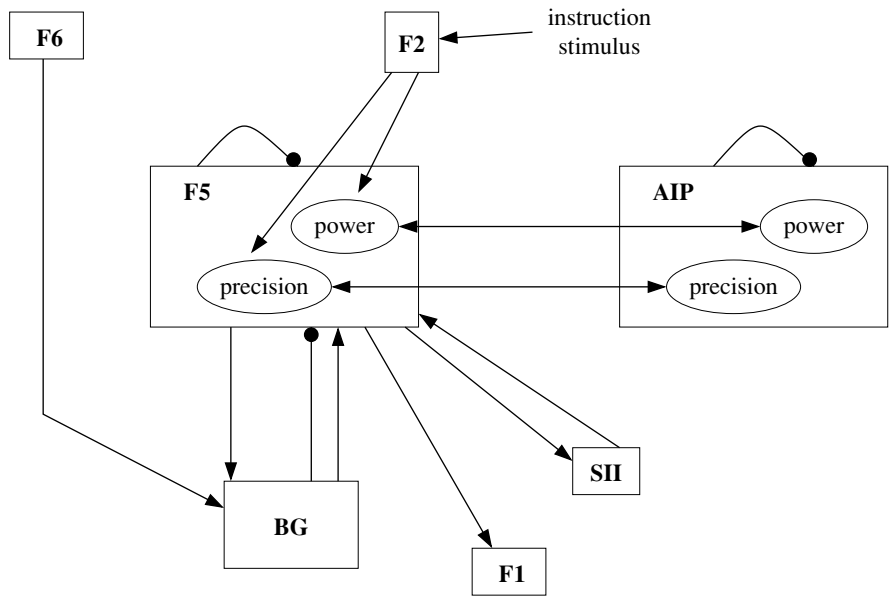


Figure 11.2 A schematic view of the model's architecture. Arrows indicate excitatory connections between regions; filled circles indicate inhibitory connections. The precision pinch and power grasp pools in F5 and AIP are connected through recurrent excitatory connections. The precision pinch pool contains more neurons than other grasps, which effects the Synthetic PET measure in these and downstream regions. F6 (pre-SMA) represents the high-level execution of the sequence, phase transitions dictated by the sequence are managed by the basal ganglia (BG). The dorsal premotor cortex (F2) biases the selection of grasp to execute as a function of the presented instruction stimulus.

phasic activation, in turn, implements the set, preshape, enclose, hold, and release phases of the grasping motor program. This monitoring of the motor program takes place at a coarse level and is not specific to the type of grasp that is executed. In general, we imagine that F6 is responsible for configuring the gating circuitry so as to implement the appropriate sequence of movements. However, this knowledge is not explicitly represented in this implementation of the model. An additional set of BG cells is responsible for providing task specific biases for the grasp selection process. These cells also receive this bias information from area F6.

11.2.2 Control of Sequential Behavior

The supplementary motor area (SMA) has been implicated in the planning and execution of complex movements (Tanji and Keisetsu, 1994). In the FARS model, area F6 (pre-SMA) is responsible for representing the high-level sequence for performing the task (wait-grasp-hold-release). This region manages the phase-related activity in F5 and F4 (ventral premotor region involved in reaching movements) via pathways through the basal ganglia (see figure 11.2). F6 first prepares the ventral premotor regions for execution of the coming grasp by priming F4 and F5. This

priming process allows set cells within F5 to become active in response to inputs from AIP. In response to the go signal given by the experimenter, F6 initiates execution of the program by priming movement-related cells in F5, and shunting set cells. Local excitatory and inhibitory interactions within F5 ensure that the selected grasp, as represented by F5 set cell activity at the time of the go signal, gives way to activation of the appropriate subpopulations of F5 movement (extension) cells. Phase transitions in F5 from extension to flexion, and from flexion to holding are triggered by either an internal model of the hand state, or directly by sensory feedback from the hand (available from SII). Initiation of the release phases of movement is also managed by F6 in response to the second go signal.

11.3 Synthetic PET for Grasp Control

In what follows, we present the results of two different Synthetic PET experiments, which serve as predictions for what we expect when the experiments are performed in the human. In both experiments, the modeled subject is asked to grasp a single object using one of two grasps.

1. In the first experiment, we examine the effects of knowing which grasp to use prior to the onset of recording (non-conditional task), as compared with only being told which grasp to use after a delay period (conditional task). In the conditional task, an instruction stimulus in the form of a bi-colored LED informs the subject which grasp should be used.

2. The second experiment looks at the differences in rCBF between a "complex" grasp (precision pinch), and a "simple" grasp (power grasp).

Table 11.1 Talairach coordinates for modeled brain regions

Brain Region	Talairach Coordinates			Source	Experimental Context
	X	Y	Z		
F5	-64	4	24	Ehrsson et al., 2000 consistent with Buccino et al., 2001	Precision vs. power grip Action observatrion
AIP	-40	-40	40	Binkofski et al., 2000 consistent with Buccino et al., 2001	Finger movement obs. Action observation
F1 complex	-26	-24	38	Kinoshita et al., 1999	Precision grip for lifting
SII	-64	-20	24	Binkofski et al., 1999	Object manipulation
F2	-31.5	- 6.1	54.2	Kurata et al., 2000	Conditional finger mov.

For both experiments, we report the relative synaptic activity for each brain region and task. In addition, the comparative activity for each brain region is painted onto a three-dimensional model of a brain, which is derived from a Talairach-

registered MRI image. Table 11.1 reports the location of each brain region and the experimental sources from which locations were derived. Regions on the surface of the model were painted if they fell within an ellipsoid surrounding the reported Talairach coordinate. The dimensions of the ellipsoids were scaled so as to bring the activation to the closest surface of the three-dimensional model.

11.3.1 Comparison of Conditional and Non-Conditional Tasks

Figure 11.3A shows the relative synaptic activity measures for the conditional and non-conditional tasks (Experiment 1). Only regions in the model that demonstrate a change in synaptic activity from one task to the other are shown.

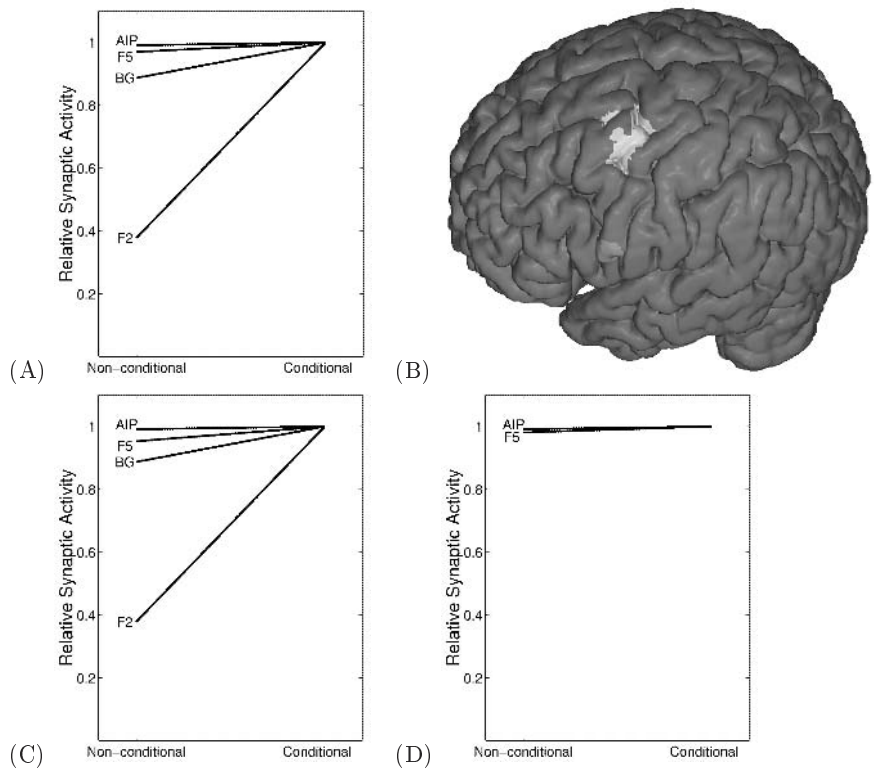


Figure 11.3 (A) Predictions of relative synaptic activity for the non-conditional and conditional tasks. Relative synaptic activity is defined as in equation 11.4. A positive slope implies an increase in relative synaptic activity from the non-conditional to the conditional task. (B) The predicted PET image that highlights the conditional synaptic activity over non-conditional activity; the deviation from gray is proportional to $rPET_A(1/2)$. (C and D) Positive and negative synapse contributions to the synaptic activity measure for the same pair of tasks.

The most significant change predicted by the model is the level of activity exhibited by area F2 (dorsal premotor cortex). Its high level of activity in the conditional task is due to the fact that this region is only involved when the model must map an arbitrary stimulus to a motor program. In the non-conditional task, the region does not receive IS inputs, and thus its synaptic activity is dominated by the general background activity in the region. The additional IS inputs in the conditional task have a second-order effect on the network, as reflected in the small changes in synaptic activity in F5, BG, and AIP. The increased synaptic activity in F5 is due to the additional positive inputs from F2. These inputs also cause an increase in the region's activity level, which is passed on through excitatory connections to both AIP and BG (recall figure 2). It is important to note that synaptic activity does not have the same meaning as neural activity. This can be seen by examining the definition of $w_{B\rightarrow A}(t)$ (see equation 2 of Section 1). The absolute value of the synaptic strength contributes positively to this measure-so increases in either positive or negative signals into a region will be reflected as an increase in synaptic activity. Neural activity, on the other hand, increases with excitatory input but decreases with inhibitory input. An important property of the Synthetic PET technique is that the positive and negative contributions to the Synthetic PET measure can be differentiated in the simulation. This information, combined with knowledge of the gross anatomy (especially the sign of connections between regions), can aid in inferring changes in neural activity across tasks. (For further discussion of the relative effects of inhibition see the Chapter "The Use of Large-scale Modeling for Interpreting Human Brain Imaging Experiments" by M.-A. Tagamets and Barry Horwitz in this volume.) Figures 11.3 C, D demonstrate the positive and negative contributions, respectively, to the overall PET measure in the conditional/non-conditional task comparison. Note that although the positive contributions to F5 and AIP essentially dominate the full PET measure, we also see small increases in the negative inputs into these regions. These inhibitory signals are due to negative inputs from local recurrent connections in the respective areas (in the case of F5, BG also contributes additional negative inputs). This serves as additional evidence that both F5 and AIP experience increases in their overall neural activity.

11.3.2 Comparison of Complex and Simple Grasps

As noted in Sec. 2.1, the model reflects the fact that the number of F5 neurons involved in the execution of the precision pinch is greater than those involved in the power grasp. We now show how this distribution is reflected in the Synthetic PET measures. The protocol used during Experiment 2 was the same as the non-conditional task described above. Figure 11.4 illustrates a general increase in synaptic activity in many of the model's regions for the precision pinch over the power grasp. This effect is due to the larger number of participating units in F5 and AIP for the precision pinch case (see figure 11.2). Not only is there a larger number of cells contributing to the rCBF measure, on average, but also each

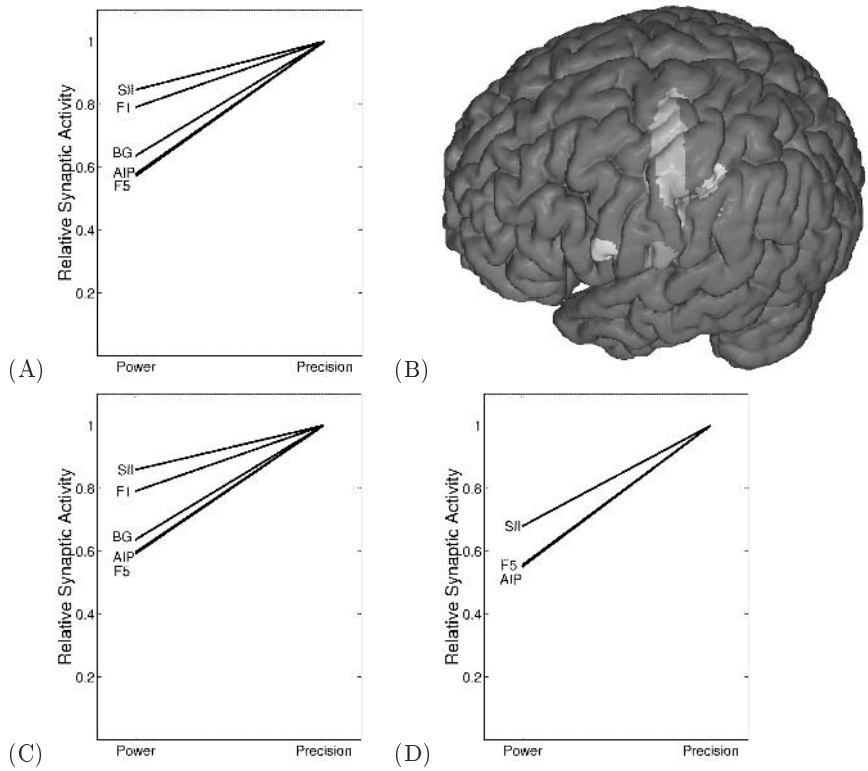


Figure 11.4 (A) Predictions of relative synaptic activity for the precision pinch and power grasp tasks. (B) Predicted PET image (precision versus power grasp). (C and D) Positive and negative synapse contributions to the synaptic activity measure the corresponding tasks.

unit in the "precision pools" of AIP and F5 receives a greater number of positive inputs from other precision units. Furthermore, despite the fact that the number of participating cells has only changed in F5 and AIP, we also observe an increase in synaptic activity in BG, F1, and SII. This is due to the fact that these regions receive input from a larger number of F5 neurons during the precision pinch task. Because these regions receive positive connections from F5 (figure 11.2), this is an indication for a true increase in *cell activity* in F5 for the precision pinch case.

11.4 Human Grasping Experiment

The Synthetic PET experiments described above raise some important questions about how instruction stimuli are mapped to arbitrary motor programs, and about the relative representation of different grasps. In this section, we summarize the results of a human PET experiment in which both of these questions were addressed

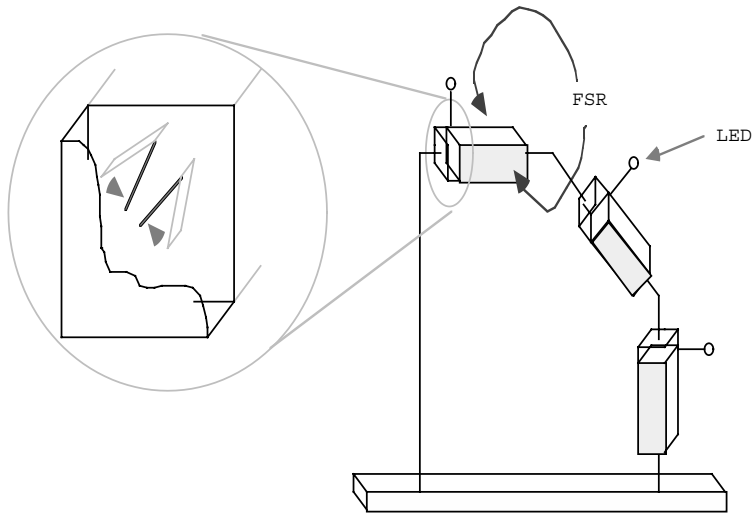


Figure 11.5 Apparatus used in PET experiment. Each of three stations can be grasped in two ways: precision pinch of the two plates in the grove (inset), or power grasp of the block. The side of the blocks are covered with a Force Sensitive Resistive (FSR) material; Light Emitting Diodes (LEDs), depending upon the task, indicate both the goal and type of grasp.

(see Grafton, Fagg and Arbib, 1998, for details of the protocol and conditional task results).

Subjects were asked to repeatedly perform grasping movements over the 90 second scanning period. The targets of grasping were mounted on the experimental apparatus shown in figure 11.5. Each of three stations mounted on the apparatus consisted of both a rectangular block that could be grasped using a power grasp, and a pair of plates (mounted in a groove on the side of the block—see inset of figure 11.5), which could be grasped using a precision pinch (thumb and index finger). A force sensitive resistive (FSR) material, mounted on the front and back of the block, detected when a solid power grasp had been established. The two plates were attached to a pair of mechanical micro-switches, which detected when a successful precision pinch had been executed. For each station, the block and plates were mounted such that the subject could grasp either one without requiring a change in wrist orientation. A bi-colored LED at each station was used to instruct the subject as to the next target of movement. A successful grasp of this next target was indicated to the subject by a change in the color of the LED. The subject then held the grasp position until the next target was given. Targets were presented every 3 +/- 0.1 seconds. Four different scanning conditions were repeated three times each. In the first, subjects repeatedly performed a power grasp to the indicated block. The target block was identified by the turning on of the associated LED (green in color). When the subject grasped the block, the color of the LED changed from green to red. For the second condition, a precision pinch was used. The target was

identified in the same manner as the first condition. In the third grasping condition (conditional task), the initial color of the LED instructed the subject to use either a precision pinch (green) or a power grasp (red). When contact was established, the LED changed to the opposite color. In the fourth (control) condition, the subjects were instructed to simply fixate on the currently lit LED, and not make movements of the arm or hand (prior to the scan, the arm was placed in a relaxed position). The lit LED changed from one position to another at the same rate and variability as in the grasping tasks. Prior to scanning, subjects were allowed to practice the tasks for several minutes.

11.4.1 Grasp versus Rest

Grafton et al. (1998) find that the areas most significantly active during grasping as compared with the non-movement (control) condition include sensory and motor areas along the central sulcus, as well as the nearby premotor and parietal cortices (figure 11.6, left), which is consistent with a number of other similar studies with arm movements (Grafton et al., 1996; Roland et al., 1982; Winstein et al., 1996). In addition, significant activity is observed in the inferior precentral gyrus/sulcus (indicated by the arrow in the left panel of figure 11.6). This region corresponds to the ventral premotor cortex, and may include the human homologue of Rizzolatti's F5 (Winstein et al., 1996).

11.4.2 Conditional versus Non-Conditional Grasp

The right panel of figure 11.6 reflects differences of conditional grasp selection (power or precision based on color cues) as compared to an average of the fixed grasping conditions (power and precision tasks): $Cond - (Power + Precision)/2$. The upper arrow indicates a large area of significance in the left superior frontal sulcus corresponding to the dorsal premotor cortex. As earlier noted, this region in monkey is thought to be involved in the arbitrary association of stimuli with the preparation of motor programs. The lower arrow indicates increased rCBF in the left inferior parietal lobule and intraparietal sulcus. Because this comparison is counterbalanced for the amount of movement made during execution of the tasks, there is no difference observed in the motor execution areas.

11.4.3 Precision versus Power Grasp

The lower panel of figure 11.6 denotes areas where rCBF activity is greater for precision grasps than for power grasps. The upper arrow indicates a site located in the left dorsal frontal gyrus, in the extreme dorsal SMA. The lower left arrow denotes a difference in the left rostral inferior parietal lobule, the lower right arrow indicates a difference in the intraparietal sulcus.

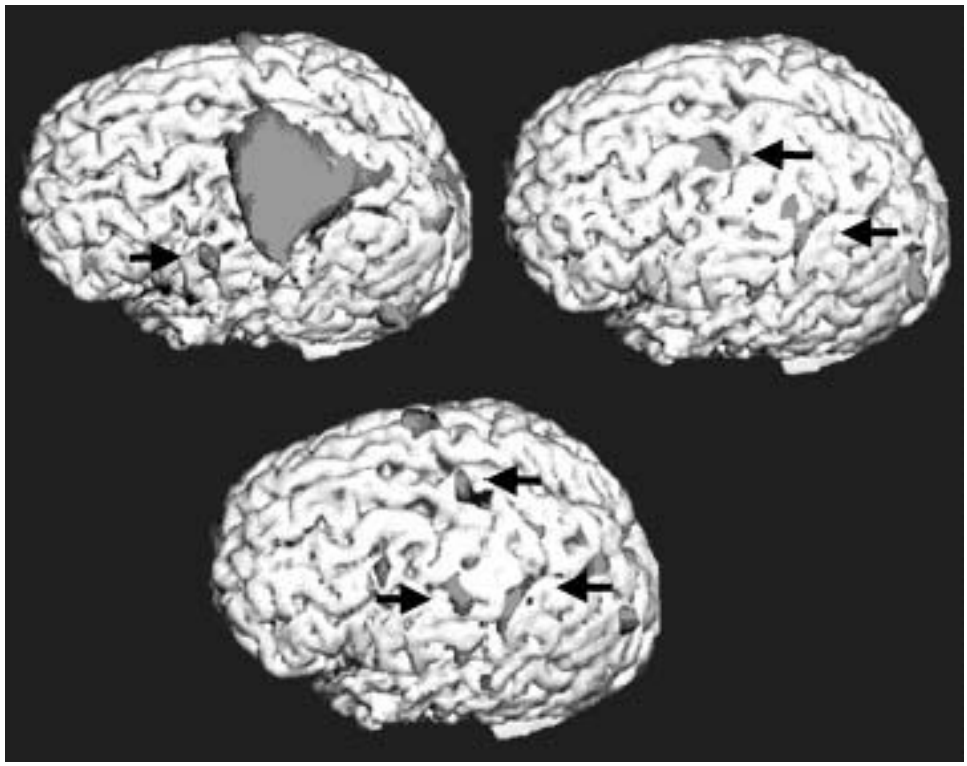


Figure 11.6 Left hemisphere localization of task related effects. PET statistical comparisons of the pooled data across subjects (darkened areas, $P < 0.005$) are superimposed on a single subject's MRI scan centered in the same coordinate space. The three panels are left superior oblique views, and denote differences of all grasp movements versus rest (left), conditional grasp selection versus fixed grasping (right), and precision versus power grasp (lower panel).

11.5 Comparison of PET and Synthetic PET for Grasp Control

11.5.1 Conditional versus Non-Conditional Task

The model predicts that the conditional task should yield much higher activation in F2 (dorsal premotor cortex), some activation of F5, and a slight activation of AIP. The human experiment confirmed the F2 result, but failed to confirm the predictions for F5. Furthermore, in human we see an activation of the inferior parietal cortex, along the intra-parietal sulcus, which is perhaps an AIP homologue.

The negative F5 result may be used to further refine the model. Consider the functional connectivity of these regions in the model (figure 11.7A). In the model, the strength of the projection from F2 to F5 is essentially a free parameter. In other words, there is a wide range of values over which the model will correctly perform the conditional and non-conditional tasks. The implication is that, by tuning this parameter, we can control this projection's contribution to the synaptic activity

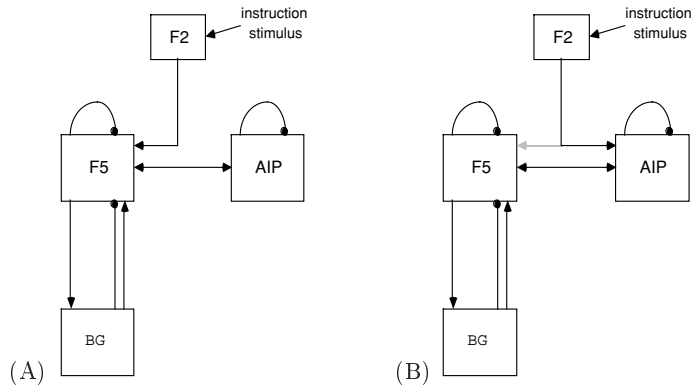


Figure 11.7 Previous functional model (A = figure 11.2) and updated functional model (B). In the revised model, the information from F2 flows (primarily) into the circuit through a projection into AIP.

measure in F5. However, the difference in AIP synaptic activity from the non-conditional to the conditional task will always be less than the difference observed in F5. This results from an interaction between the neural dynamics and the assumptions made about the model anatomy. Suppose that the projection strength from F2 to F5 is increased. In this case, we would observe an increase in both F5 synaptic and cell activity. The increase in F5 cell activity, however, is attenuated by local, recurrent inhibitory connections. Thus the excitation that is then passed on to AIP via F5 does not reflect the full magnitude of the signal received from F2.

The conclusion is that, although we can adjust the free parameter to match one or the other observations in the human experiment (of either F5 or AIP changes), the model cannot reflect both at the same time. One possibility for repairing this problem in the model is to reroute the F2 information so that it enters the grasp decision circuitry through AIP (or both AIP and F5), rather than exclusively through F5 (figure 11.7B). This would yield an increase in activity in AIP due to F2 activation with only an attenuated signal being passed on to F5, resulting in only a small increase in F5 synaptic activity. Note that we do not necessarily assume that there exists a direct cortico-cortical connection from F2 to AIP or F5, but only that there is a functional connection (which potentially involves multiple synapses).

11.5.2 Precision versus Power Task

The model predicts a higher degree of synaptic activity in both F5 and AIP for complex grasps than for simple grasps (in the model, these grasps are the precision pinch and power grasp, respectively). Although we see an increase in activity for the precision grasp case along the inferior parietal sulcus (IPS) in the human experiment, we fail to see any such change in the ventral premotor

cortex (specifically F5). Two explanations are possible for this negative result. First, although Rizzolatti et al. (1988) observe a larger number of cells for the more complex grasps (e.g. the class of precision pinches), these cells may be involved in encoding the many variations of the grasp. But, when a specific grasp is executed, the number of active cells may be the same as for any other grasp instantiation (including side oppositions and power grasps). If this is the case, however, the model would also predict no difference in synaptic activity in AIP. The second possibility assumes that the number of active F5 cells does indeed differ significantly between the precision and power grasps, but that the effect is masked by force-related activity in the region. In the human experiment, performance of the power grasp required a reasonable level of force to be applied to the block before the LED would indicate to the subject that a grasp had been detected. In monkey, force-related activity has been observed in F5 (Hepp-Reymond et al., 1994). The implication is that even though there are fewer neurons involved in encoding the power grasp, they achieve a higher level of activity because of the force requirements of the task. This higher level of cell activity is an indicator of increases in the cells' inputs, which implies an increase in the rCBF measure. Thus, the rCBF measures could be similar enough in the two conditions to not be detectable above the noise levels inherent in the PET imaging process.

11.6 Discussion

The fundamental benefit of the Synthetic PET method is that it allows for specific predictions in PET experiments based on neural network models of behavior. Since the models themselves are a product of functional anatomy, measured single-unit recordings, and behavioral measurements, Synthetic PET provides a powerful bridge between all of these approaches. An additional strength of the Synthetic PET implementation is that the contribution of excitatory and inhibitory influences can be teased apart. Because synaptic activity is not the same as neural activity, being able to distinguish excitatory from inhibitory influences can be an aid to inferring neural activity from the rCBF measure, possibly clarifying apparent contradictions in rCBF data (an example has been demonstrated in Arbib et al. (1995). The low-level details of the FARS grasping model (Fagg and Arbib, 1998) were derived primarily from neurophysiological results obtained in monkey. The Synthetic PET approach extracts measures of regional synaptic activity as the model performs a variety of tasks. These measures are then compared to rCBF (regional cerebral blood flow) observed during human PET experiments as the subjects perform tasks similar to those simulated in the model. In some cases, the human results provide confirmation of the model behavior. In other cases, where there is a mismatch between model prediction and human results, it is possible (as we have shown) to use these negative results to further refine and constrain the model and, on this basis, design new experiments for both primate neurophysiology and human brain imaging. An additional feature of the Synthetic PET technique is that it provides

a link between neural network models and anatomic circuitry with the results displayed directly on a brain atlas centered in Talairach coordinates (Talairach and Tournoux, 1988). This facilitates interaction between anatomists, physiologists and modelers interested in common neurobehavioral phenomena. The method is sufficiently flexible that it will be possible to have network implementations spanning multiple species. Homologies and differences between species can then be tested more rigorously using predictions generated by the Synthetic PET , while the human data provide another form of validation of neural network models derived from monkey data.

Our current measure of "raw PET activity", based on a linear function of the total of the absolute value of synaptic activity, already yields qualitatively useful results in evaluating the sign and small versus large magnitude of activities seen in PET comparisons. However, we do not claim that this first approximation yields quantitatively accurate predictions. We note, as a target for further research on Synthetic brain imaging, the interest of evaluating a variety of more quantitative fits based on (possibly nonlinear) combinations of cell firing rates, synaptic change, and synaptic activity per se. We also note the possibility of performing a stochastic analysis with the model in order to account for the variation in PET activity seen in the same subject over a set of trials.

Acknowledgments

This work was supported by Public Health Service grants NS 33504 and NS01568 (S.G.) and a Human Frontier Science Program grant (S.G., M.A.). The authors wish to thank Roger P. Woods M.D. of the University of California at Los Angeles for generous software support as well as Kim Hawley and Steve Hayles for technical assistance.

References

M. A. Arbib. Part I: The Background. In M. A. Arbib (Eds.), The Handbook of Brain Theory and Neural Networks (pp. 3-25). The MIT Press, Cambridge, MA, 1995.

M. A. Arbib, A. Billard, M. Iacoboni, and E. Oztop. Synthetic Brain Imaging: Grasping, Mirror Neurons and Imitation, Neural Networks, 13, 975-997, 2000.

M. A. Arbib, A. Bischoff, A. H. Fagg, and S. T. Grafton. Synthetic PET: Analyzing Large-Scale Properties of Neural Networks. Human Brain Mapping, 2, 225-33, 1995.

F. Binkofski, K. Amunts, K. M. Stephan, S. Posse, T. Schormann, H.-J. Freund, K. Zilles, and J. Seitz. Broca's Region Subserves Imagery of Motion: A Combined Cytoarchitectonic and fMRI Study. Human Brain Mapping, 11, 273-85, 2000.

F. Binkofski, G. Buccino, S. Posse, R. J. Seitz, G. Rizzolatti, and H.-J. Freund. A Fronto-Parietal Circuit for Object Manipulation in Man: Evidence from an fMRI-study. European Journal of Neuroscience, 11, 3276-86, 1999.

M. Bota and M. A. Arbib. The NeuroHomology Database, in Computing the Brain: A Guide to Neuroinformatics, (M. A. Arbib, and J. S. Grethe, Eds.), San Diego: Academic Press, pp.337-351, 2001.

G. L. Brownell, T. F. Budinger, P. C. Lauterbur, and P. L. McGeer. Positron Tomography and Nuclear Magnetic Resonance Imaging. Science, 215, 619-26, 1982.

G. Buccino, F. Binkofski, G. R. Fink, L. Fadiga, L. Fogassi, V. Gallese, R. J. Seitz, K. Zilles, G. Rizzolatti, and H.-J. Freund. Action Observation Activates Premotor and Parietal Areas in a Somatotopic Manner: an fMRI Study. European Journal of Neuroscience, 13, 400-4, 2001.

H. H. Ehrsson, A. Fabergren, T. Jonsson, G. Westling, R. S. Johansson, and H. Forssberg. Cortical Activity in Precision- Versus Power-Grip Tasks: An fMRI Study, Journal of Neurophysiology, 83(1), 528-36, 2000.

E. V. Evarts, Y. Shinoda, and S. P. Wise. Neurophysiological Approaches to Higher Brain Function. Wiley-Interscience Press, New York, NY, 1984.

A. H. Fagg. A Computational Model of the Cortical Mechanisms Involved in Primate Grasping. Ph.D. Dissertation, University of Southern California, Computer Science Department, 1996. Available at: http://www.cs.umass.edu/∼fagg

A. H. Fagg and M. A. Arbib. A Model of Primate Visual-Motor Conditional Learning. Journal of Adaptive Behavior, 1(1), 3-37, 1992.

A. H. Fagg and M. A. Arbib. Modeling Parietal-Premotor Interactions in Primate Control of Grasping. Neural Networks, 11(7/8), 1277-1303, 1998.

P. T. Fox and M. E. Raichle. Stimulus Rate Determines Regional Brain Flow In Striate Cortex. Annals of Neurology, 17, 303-5, 1985.

Grafton, S. T., Fagg, A. H., and Arbib, M. A. (1998). Dorsal Premotor Cortex and Conditional Movement Selection: A PET Functional Mapping Study. Journal of Neurophysiology, 79(2), 1092-7.

S. T. Grafton, A. H. Fagg, R. P. Woods, and M. A. Arbib. Functional Anatomy of Pointing and Grasping in Humans. Cerebral Cortex, 6(2), 226-37, 1996.

M. C. Hepp-Reymond, E. J. Husler, M. A. Maier, and H. X. Qi. Force-Related Neuronal-Activity in 2 Regions of the Primate Ventral Premotor Cortex. Canadian Journal of Physiology and Pharmacology, 72(5), 571-9, 1994.

H. Kinoshita, N. Oku, K. Hashikawa, and T. Nishimura. Functional Brain Areas Used for the Lifting of Objects Using a Precision Grip: a PET study. Brain Research, 857, 119-30, 2000.

K. Kurata, and S. P. Wise. Premotor Cortex of Rhesus Monkeys: Set-Related Activity During Two Conditional Motor Tasks. Experimental Brain Research, 69(2), 327-43, 1988.

K. Kurata, T. Tsuji, S. Naraki, M. Seino, and Y. Abe. Activation of the Dorsal Premotor Cortex and Pre-Supplementary Motor Area of Humans During an Auditory Conditional Motor Task, Journal of Neurophysiology, 84(3), 1667-72, 2000.

A. R. Mitz, M. Godshalk, and S. P. Wise. Learning-Dependent Neuronal Activity in the Premotor Cortex. Journal of Neuroscience, 11(6), 1855-72, 1991.

W. Rall. Perspective on Neuron Model Complexity. In M. A. Arbib (Eds.), The Handbook of Brain Theory and Neural Networks (pp. 728-732). MIT Press, Cambridge, MA, 1995.

G. Rizzolatti, R. Camarda, L. Fogassi, M. Gentilucci, G. Luppino, and M. Matelli. Functional Organization of Inferior Area 6 in the Macaque Monkey II. Area F5 and the Control of Distal Movements. Experimental Brain Research, 71, 491-507, 1988.

P. E. Roland, E. Meyer, T. Shibasaki, Y. L. Yamamoto, and C. J. Thompson. Regional cerebral blood flow changes in cortex and basal ganglia during voluntary movements in normal human volunteers. Journal of Neurophysiology, 48, 467-80, 1982.

M. Taira, S. Mine, A. P. Georgopoulos, A. Murata, and H. Sakata. Parietal Cortex Neurons of the Monkey Related to the Visual Guidance of Hand Movement. Experimental Brain Research, 83, 29-36, 1990.

J. Talairach and P. Tournoux. Co-planar Stereotaxic Atlas of the Brain. Thieme Medical Publishers, New York, 1988.

J. Tanji and S. Keisetsu. Role for Supplementary Motor Area Cells in Planning Several Movements Ahead. Nature, 371(29), 413-6, 1994.

C. J. Winstein, S. T. Grafton, and P. S. Pohl. Motor task difficulty and brain activity: an investigation of goal-directed reciprocal aiming using positron emission tomography (PET). Journal of Neurophysiology, 77, 1581-94, 1996.

S. P. Wise, and K. H. Mauritz. Set-related Neuronal Activity in the Premotor Cortex of Rhesus Monkeys: Effects of Changes in Motor Set. In Proceedings of the Royal Society of London, 223, 331-54, 1985.

12 The Use of Large-Scale Modeling for Interpreting Human Brain Imaging Experiments

M.-A. Tagamets
Maryland Psychiatric Research Center
University of Maryland School of Medicine
Baltimore, Maryland 21228 USA
mtagamet@mprc.umaryland.edu

Barry Horwitz
Voice, Speech, and Language Branch
National Institutes of Health
Bethesda, MD 20892 USA
horwitz@helix.nih.gov

Functional neuroimaging methods such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have become increasingly important in the study of human brain function. These methods allow the study of the conscious, behaving human brain, and are commonly used for relating regional brain activations to specific cognitive components. However, there is still a large gap between the imaging results and underlying neuronal processes such as those described in single-cell animal studies. A few efforts have been made to bridge this gap by the use of large-scale modeling. Some of the factors and constraints that need to be considered in building such a model are discussed. Then a specific model of visual working memory is described. The model performs a task that has been used in both animal single-cell recordings and human brain imaging experiments. The model includes elements that have dynamics similar to the various neuronal populations that have been identified in the ventral visual pathway, while at the same time the total summed synaptic activity in the different regions is similar to human imaging data. In this model, the emphasis is at the circuit level, and the expected effect of balance of excitatory and inhibitory connections on imaging data is discussed. Specific experiments and predictions made by the model include

the expected effect of synaptic inhibition on imaging data, an interpretation of transcranial magnetic stimulation (TMS) data, and a potential mechanism for the mediation of working memory in the prefrontal cortex.

12.1 Introduction

Human brain imaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have come to be important in relating human cognition to physical changes in the brain. There is a great potential for these studies to gain from other areas of neuroscience research, such as single-cell recordings in animals, anatomical and circuit level structural studies, and physiological studies of systems such as synaptic transmitter systems. These latter types of studies can potentially inform the interpretation of the human imaging data and enhance our understanding of human neurophysiology. However, because these methods all yield different measures of neuronal activity at different temporal and spatial scales, interpreting the data into a coherent and unified picture has proved difficult.

A better understanding of the relationships between these widely differing methods requires first identifying salient aspects of how each of these measures reflects the underlying neuronal activity. This allows a formulation of mechanisms that "translate" between them. Large-scale modeling is one tool that is ideally suited for this purpose. We have developed such a model (Tagamets and Horwitz, 1998) with the specific goal of understanding the neural underpinnings of human imaging data and relating it to animal studies. The model performs a delayed match-to-sample (DMS) task, a paradigm that has been used in both animal single-cell recordings and human brain imaging studies. It includes four separate regions that model the ventral visual pathway, and incorporates a circuit in the model frontal cortex area that maintains delay-period activity, acting as a short-term memory. Quantitatively, the simulated regional blood flow matches that of human imaging experiments of the DMS task. Qualitatively, the temporal behavior of the individual units in the model are similar to those observed in animal single-cell recordings. In contrast to other large-scale models of working memory, the design is based on specific factors that are likely to influence how neuronal activity is manifested quantitatively as the data seen in human imaging. In this paper, we have focused on three major classes of information that are expected to be important in affecting this relationship: 1) The direct relationship between spiking and fMRI and PET at a cellular level; 2) The influence of local and inter-regional circuit properties; 3) Modulatory effects, e.g. from other brain regions or from neurotransmitters such as dopamine. A number of other issues relevant to models of imaging are reviewed in Horwitz and Sporns (1994).

12.2 Relating Spiking Activity to Brain Imaging Data

In studies of information coding at the single-cell electrophysiological level, the most commonly measured entity is the rate of action potentials (*spiking*) of single neurons. In general, the rate of spiking is taken to be the de facto standard for assessing the coding properties of neurons, though there have been a number of studies that examine information encoded by timing relationships among either separate neurons or populations of cells (Gray et al., 1989; Aersten et al., 1989; Eskandar et al., 1992; Friston, 1997). While single-cell recordings provide a direct measure of this spiking activity, the human imaging methods of PET and fMRI are indirect measures, and apply to large populations of neurons, having a spatial resolution of about 1 cm^2 of the brain's surface, which includes about 10 to 30 million neurons (Peters, 1987). It has been estimated that anywhere from 60 to 80% of these neurons are pyramidal cells, which are those nerve cells whose spiking properties are most often studied by single-cell methods. Specifically, regional cerebral blood flow (rCBF, measured by PET) and blood oxygenation level (BOLD, measured by fMRI) are thought to reflect local energy requirements that are a byproduct of synaptic activity. Whereas rCBF is the measured blood flow in a brain region, BOLD reflects the oxygenation level in the blood. Changes in the latter, measured by fMRI, are caused by an increase in oxygenated blood that is carried in by increased blood flow. In summary, a single voxel in PET and fMRI data represents a transformed, mixed blend of neural activities that vary both in their coding properties and changes in their background state over time.

To date, most imaging studies examine *relative* changes in brain activity in a subtraction paradigm, yielding "hot spots" of activation in circumscribed regions that are interpreted to be important in one task relative to another. It is thought that such changes in rCBF and BOLD measure energy demands that are needed for recycling transmitters at presynaptic terminals (for a review see Jueptner and Weiller, 1995). Consequently, it is likely that both excitatory and inhibitory synaptic activities can cause a rise in rCBF or BOLD activity. More recently, it has been suggested that astrocytes play a crucial role in this cycle at the excitatory glutamatergic synapses (Magistretti and Pellerin, 1999). Two factors are especially germane to modeling the relationship between human imaging data and single cell recordings: 1) glutamatergic synapses are the most abundant in the cortex, and (2) the glutamate cycle as mediated by astrocytes is tightly coupled to changes in local metabolism that are generated from the activity of neurons (Magistretti and Pellerin, 1999). This suggests that there is a direct relationship between excitatory synaptic activity and energy metabolism, and this has indeed been shown to be true in circuits that have known excitatory inputs (Sokoloff, 1993). However, it has been proposed that energy demands at inhibitory synapses also would tend to increase blood flow, even when local spiking activity decreases, and results suggesting this have been reported in known inhibitory pathways (Ackermann et al., 1984; Mathiesen et al., 1998). Thus, in a patch of cortex at the resolution of

PET and fMRI, the corresponding rCBF or BOLD activation reflects a blend of both excitatory and inhibitory local and afferent synaptic activity. A consequence of this is that it may not be apparent whether PET or fMRI activations reflect increased or decreased spiking, especially if there is a mix of synaptic types in the area being measured. Since it is known that inhibitory interneurons and inhibitory pathways play a crucial role in the regulation of brain activity, this is an important factor in understanding human neurophysiology in terms of brain imaging. This also leads to the basic premise that imaging data can be modeled as the sum of all synaptic activities (both excitatory and inhibitory) in a model area that corresponds to the region of interest. This method was used by Arbib et al. in a model of memory-driven saccades (Arbib et al., 1995) and in a study of visually directed hand movements (Arbib et al., this volume). We have also taken this approach to representing rCBF/BOLD activations in our model (Tagamets and Horwitz, 1998; Horwitz and Tagamets, 1999).

Because of the mix of excitatory and inhibitory signals that make up a unit of imaging data, the type and strength of connectivity is likely to play a major role in relating spiking to imaging data. Local connectivity in particular is thought to dominate in the cerebral cortex, and the role of this in the model will be discussed later. Functional connectivity between regions has recently become a topic of great interest in the field of neuroimaging (Horwitz et al., 1992; Friston, 1994; Horwitz, 1994; McIntosh et al., 1996; Friston et al., 1997). Taylor et al. (Taylor et al., 2000) derived theoretical conditions for making valid inferences about functional connectivity from structural equation modeling. The two main results of their work suggest that from the point of view of imaging data, all neuronal (spiking) activity is a hidden variable, and that it is necessary to account for both excitatory and inhibitory neuronal populations when interpreting the results. One of the strengths of modeling is that the relative contributions of each of these populations can be observed independently. We have used our model to examine how different types of inter-regional inputs can affect both spiking and simulated blood flow, and demonstrated conditions under which there is a dissociation between the two (Tagamets and Horwitz, 1997; Tagamets and Horwitz, 2001). This will be reviewed in more detail in section 12.4.

Another factor that affects rCBF and BOLD is the background activity of neurons in the region being examined. In a combined experimental and theoretical study, Scannell and Young (1999) presented results that suggested that changes in the background activity (without a change in spiking level) could modify measured rCBF at least as much as changes in spiking alone. This highlights the importance of modulatory effects in imaging. In particular, there is an interaction of sensory input (feed-forward) and modulatory influences (feedback or transmitter-based) in most regions of the cortex, especially in association areas. In a model, these effects can be explicitly specified and they can be systematically examined. For example, different parameters in the dynamics of model units can be used to implement the effects of neuromodulators such as dopamine, which are thought to play a major role in working memory. In section 12.5, we show how we have used this information

to improve performance of the model in the task and, at the same time, better approximate observed experimental imaging data in frontal cortex (Tagamets and Horwitz, 2000). The next section reviews the structural and functional specifications of the model. Examples of simulation studies of inhibitory effects and working memory modulation follow.

12.3 The Model

Our model was specifically developed for simulation of imaging data. The model performs a visual delayed match-to-sample (DMS) task, electrical activity of model elements are similar to those found in primate electrophysiology data in a DMS task, and simulated PET activity is similar to human PET data in the same task. One of the main goals for the model was to include anatomical and physiological constraints that are likely to influence the relationship between the imaging and single-cell methods. This led us to pay particular attention to the properties of a local circuit that forms the basic element of the model and to the nature of inter-regional connections. It is well known that local connections are abundant in the cerebral cortex, and that there is a great deal of local recurrent feedback (Douglas et al., 1995). Activity at a site also depends on the type and number of afferent synapses from other regions.

The goal in designing the model was to take into account the balance of excitatory and inhibitory synapses in both local and inter-regional circuits. The local basic circuit was defined first, and was designed to approximate known proportions of local excitatory and inhibitory connections in the cortex. It is made up of two units, which represent the local excitatory population (E) and the local inhibitory population (I) of neurons in a circuit that approximates a cortical column (see figure 1A). Based on animal studies (Douglas et al., 1995), the proportions of four types of connections were evaluated and included in the model: 1) Local **E** $\rightarrow$ **E:** 60% of all connections; 2) Local **E** $\rightarrow$ **I:** 15% of connections; 3) Local **I** $\rightarrow$ **E:** 15% of connections; and 4) Afferent connections: the remaining 10%. The basic unit of the model captures these properties (see figure 1A), and the local connection proportions are identical in all regions of the model. A patch of cortex is modeled by a 9x9 group of local basic circuits, as shown in figure 1B, and represents about 1 cm^2 of cortex. Electrical activity that models the spiking of neurons is computed by a sigmoidal activation rule, with parameters fixed in all elements of the model. rCBF in the model is computed as the sum the absolute values of all synaptic activations, both excitatory and inhibitory. For more details on the derivation of the connections and parameters, see Tagamets and Horwitz (1998).

Differences in inter-regional connectivity account for different behaviors in the regions of the model, while the interactions of local and inter-regional connections determine simulated blood flow in each region. The full-size model is created by connecting 9x9 patches of local units together (see figure 2). Regions represented in the model are primary visual cortex (V1/V2), ventral extrastriate cortex (V4),

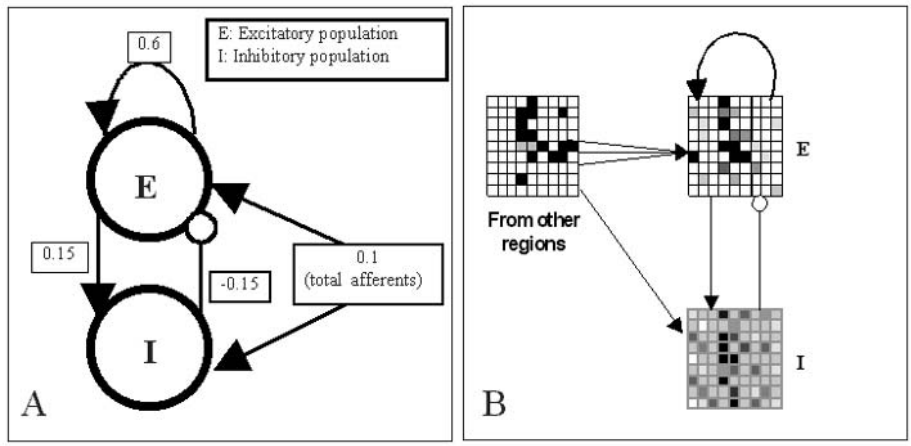


Figure 12.1 (A) The local circuit consists of one excitatory unit (E) and one inhibitory unit (I) that are connected as shown. 60% of connections are local and excitatory-to-excitatory, 15% are local and excitatory-to-inhibitory, and 15% are local and inhibitory-to-excitatory. 10-20% of all connections are afferents from other regions. (B) A patch of cortex is made up of a 9x9 array of local circuits as depicted in (A). Each square represents a single unit in the patch. E and I elements are shown separately, but are connected as in (A).

anterior fusiform and inferior temporal cortex (TEO/IT) and prefrontal cortex (PF). Between-region connections are described by 3 parameters and then generated automatically. The parameters are: 1) The fanout size in each of the two dimensions (mxn); 2) The sparseness (percent of connections that actually exist in the mxn area); and 3) The total weight (i.e. the sum of all connection weights that arise from a single unit and converge on the mxn area). Thus it is possible to examine how each of these parameters can influence both spiking and rCBF. For example, given the same total weight and fanout, increasing the sparseness makes individual units more selective, while total rCBF in the recipient region goes down. The reason that rCBF goes down is that fewer neurons respond in the local region, even though those that do respond are more selective and have increased spiking rates. In animal studies, it is usually the selectivity of single neurons that is reported in the literature. Understanding how connectivity affects both coding properties and imaging data simultaneously will help lead to a better understanding of human cortical mechanisms.

12.4 Inhibition Effects in a Single Region

Some aspects of the relationship between imaging and neuronal data can be examined in a single region of the model. The observed hemodynamic response is the result of an interaction between afferent activity and local response. The basic element takes into account the relative importance that local circuits are thought to

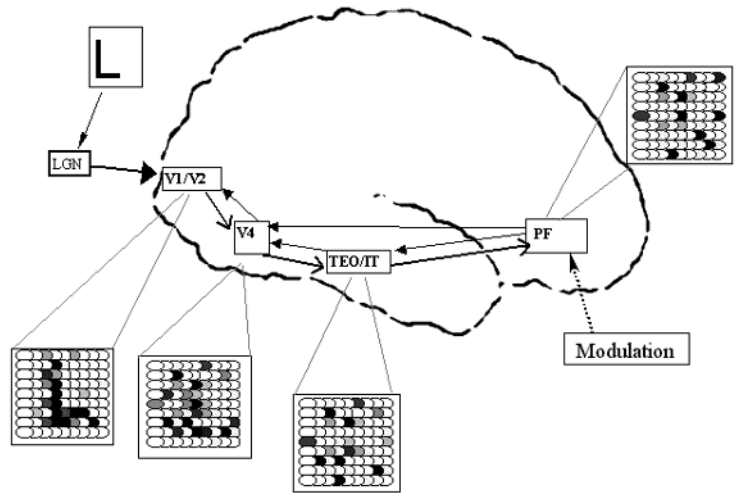


Figure 12.2 The large-scale model is composed of connected patches (see figure 1B). Each of the regions has separate subgroupings of units that have selective response properties. For example, area V1/V2 has two populations, one which is selective for horizontal lines and one of vertical lines. Simulated imaging data is computed as the average synaptic activity across all populations within each region. The frontal region is made of specialized memory modules that implement the working memory circuit (see section 12.5). Stimuli are presented to the lateral geniculate nucleus (LGN) region, and are made up of simple geometric figures composed of straight lines. Patches of units shown in the expanded views depict the increasing abstraction and receptive fields as activity moves forward along the pathway from V1/V2 to frontal cortex (PF).

have in shaping a neuronal response within a local cortical region. Even in a single unit of the model, there is a nontrivial relationship between mean firing rate of the excitatory units in an area (i.e. pyramidal cells, those usually recorded in non-human electrophysiological studies) and total synaptic activity in the population of all cells. This is further complicated by the fact that within a region, there is a mixture of elements that have different response profiles for different stimuli. One open question concerns the relationship between afferent inhibition, spiking activity and imaging data.

In order to examine this question, we used a single model area **A.** The effects of two factors were examined: 1) the amount of local recurrent excitation, and 2) the amount of excitation that is driving the circuit from elsewhere (e.g. as during a task and a passive control condition). The total amount of recurrent local connectivity was set at three levels, as determined by the connection strengths of the E→E connections: 1) none, 2) medium recurrence (0.6, as in the basic circuit of the model, figure 1A), and 3) high recurrence (E→E connection strength = 0.9). In the absence of any local recurrent excitatory connectivity, most synaptic activity is from an external source. In this case, any afferent inhibition could potentially reduce spiking and raise overall synaptic activity. In contrast, at very high levels

of excitatory recurrence, it is predicted that the strong nonlinearities induced by this component would dominate in determining total synaptic activity. In this case, rCBF would decrease if spiking decreases and vice versa. In a local circuit such as shown in figure 1A, with a moderate amount of excitatory recurrence, it is not obvious what the net effect of inhibition would be. It might be expected that this would depend to a large extent on the background activity of the circuit. For example, if there is very low background, any new incoming inhibition contributes a substantial proportion of local synaptic activity, and rCBF might be expected to rise. On the other hand, if the circuit is being driven by afferent excitation, the nonlinear effects dominate the total amount of local synaptic activity, and rCBF would go down concurrently with spiking. Local spiking goes down in both cases, but there is a dissociation in the rCBF.

We examined how inhibition affects rCBF in a model area **A** under three levels of local excitatory-to-excitatory connection strengths, and at two different levels of afferent excitation that would be similar to a task condition and a control condition. Simulations were performed by presenting a pattern to a fixed array of excitatory inputs that project to the excitatory population of the area and computing rCBF for the duration of the stimulus. Two types of background activation patterns are used: (1) afferent excitatory input is applied to a subset of the E units of region A; this depicts an active task, in which the region is being driven from other sources; and (2) all inputs are set to a uniformly low value to simulate a low level of afferent input to the region. Each task is run once with added inhibition and once without. Inhibition is implemented as a low level of afferent excitation to all of the local inhibitory units in area **A.** This is the configuration that was experimentally examined by Mathiesen et al. (Mathiesen et al., 1998), in which they found that afferent inhibition caused spiking decreases but blood flow increases.

Figure 3 shows the results of the simulations for each of the conditions. Up-arrows signify increased rCBF with increasing inhibition, and down-arrows the opposite. In each case, spiking activity goes down. Thus, conditions where rCBF increases (up-arrows) indicate a dissociation between spiking and rCBF activity. Of particular note is the center row, in which local recurrence is at the level thought to exist in cortical circuits. In this case, the existence of dissociation depends on the context in which inhibition occurs. At a low background excitation level, analogous to a condition in which there is no afferent excitation from other regions, there is a dissociation between spiking and rCBF, while with high background activity, both spiking and rCBF go down. These results suggest that inhibition may either lower or raise the hemodynamic response in a cortical region, depending both on the amount of local excitatory feedback in the region, and on whether the site is being activated from other sources.

In summary, at the moderate to high level of recurrence that is thought to exist in local circuits in the cortex, afferent inhibition may cause a dissociation between spiking and rCBF change that is further modulated by the activity context in which the inhibition occurs. Specifically, even though spiking goes down, rCBF may go either up or down, depending on how much other activity is present in the circuit.

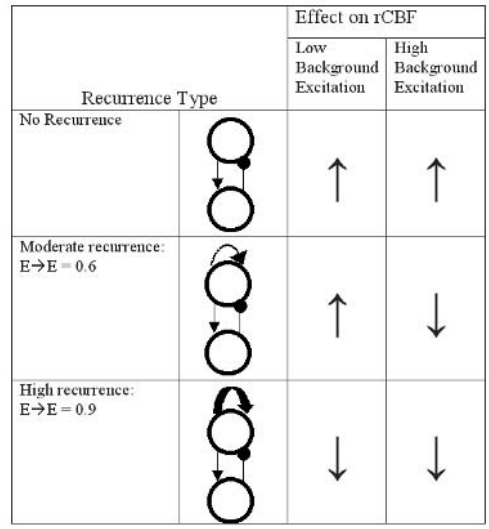


Figure 12.3 Effects of inhibition on modeled rCBF. Up-arrows represent cases where rCBF increases, and down-arrows where it decreases. Each row shows the effects at a different degree of local excitatory recurrence, with the middle row depicting the local circuit as used in the mode, and is the amount of recurrence thought to exist in the cortex. The two columns on the right show rCBF changes in each of two condition: 1) with low background activity in the recipient region, and 2) with a high amount of afferent excitation arriving in the region concurrently with the inhibition. In all cases, spiking activity within the region goes down, so that condition where rCBF goes up are those in which there is a dissociation between spiking and rCBF.

Such effects need to be explored in more detail before definitive interpretation of the neuronal dynamics underlying PET and fMRI data is possible. These considerations are especially relevant to the understanding of imaging results from patient populations, since the effects of focal lesions, in which whole populations of columns are removed, are likely to be quite different from those that result from more diffuse degenerative processes, where local connectivity is reduced, or from disconnection between areas, in which local circuitry may remain unchanged but total afferents are reduced.

12.5 Working Memory and Modulation

Multiple regions of the model can be connected to form a large-scale network that represents a set of connected regions in the cortex. Simulating an entire human imaging study calls for a model that is able to perform several different tasks, so that the results from one task can be subtracted from the other to yield quantitative data that approximate a subtraction paradigm in imaging. Furthermore, the timing and phases of the tasks should emulate those from the imaging study, since imaging data are typically acquired over a period of seconds to minutes, during which there

may occur a sequence of several different cognitive components that are averaged together in time. The frontal region of the model includes a local memory circuit, which is the main component that performs the working memory task. In this section we illustrate how a PET subtraction study can be modeled by running simulations that correspond to two different conditions, a task condition and a control condition.

The large-scale model includes areas that represent four different cortical areas: V1/V2, V4, TEO/IT, and lateral frontal cortex (PF). These regions are thought to comprise the major object processing pathway (Ungerleider and Mishkin, 1982; Haxby et al., 1991; Haxby et al., 1994), and there is extensive animal single-cell data for each of these regions in a working memory task (Fuster et al., 1982; Haenny et al., 1988; Fuster, 1990; Wilson et al., 1993; Miller et al., 1996). Although it is not known exactly what the homology is between humans and non-human primates in these areas, anatomical and imaging studies suggest that at least the earlier parts of this pathway are similar between the two species (Burkhalter and Bernardo, 1989.; Sereno et al., 1995.; DeYoe et al., 1996). The connections between the regions of the model are constructed with two constraints in mind: 1) Spiking activity should reflect encoding properties that are similar to neurons found in each of the regions; and 2) The total inter-regional connection strength should approximate known proportions of between-region connectivity. Given constraint 2, connections with a fanout and topographic pattern are generated in order to meet constraint 1. This yields appropriate selective response properties for each region: units in V1/V2 respond optimally to short line segments; units of V4 have optimal responses to longer lines and corners, and those in TEO/IT are selective for whole simple geometric figures that are composed of straight lines. The object-specific tuning in the IT units was obtained by a competitive Hebbian learning rule. For more details on the construction of the model, see Tagamets and Horwitz (1998).

The frontal region of the model has four different types of units, whose responses mediate the online storage, retrieval and decision phases of working memory (see figure 4). The behaviors of these units are based on results from single-cell recordings in the monkey during a delayed response task, and are defined in terms of the time when they are active, i.e. during the cue period, the delay period, or the response period of the task (Funahashi et al., 1990; Goldman-Rakic, 1995). Cue sensitive cells (**C**) respond only when a stimulus is present and response neurons (**R**) display a brief response on presentation of the second stimulus if it matches the one in memory. Delay-active neurons in the prefrontal cortex have been classified by Funahashi et al. (1990) as being of two types: those that are active only during the delay period (**D1** unit type), and others that display activity during both the cue and the delay periods (**D2** unit type). Performance of the task is mediated by a gating mechanism at the D2 units. When gating is ON, a "high attention" state is modeled, and the circuit performs the DMS task by maintining a representation of the stimulus over the delay period via a high activity level in the D1 and D2 units. With gating turned OFF, no memory trace is maintained, and the circuit is unable to perform the task. The specific form of gating is described later in this

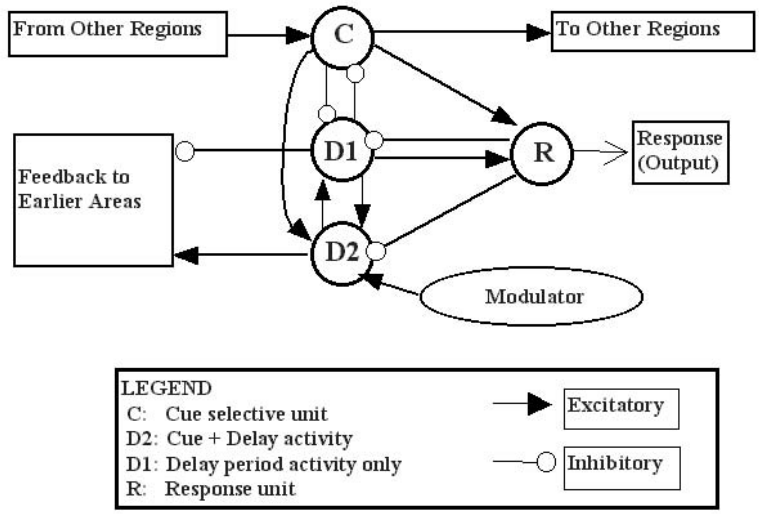


Figure 12.4 The working memory circuit in the model frontal cortex area has four types of units with different temporal activity profiles. C: Cue units, active only when an external stimulus is present; D1 and D2: Delay-active units. These maintain the representation of items to be remembered in an excitatory feedback loop. D2 units are active both during the delay period and when a stimulus is present, and are also the recipients of the gating signal, which controls updating and maintenance of items in memory. D1 units become active only when the external stimulus is not present and remain active during the delay period. R: Response units mediate the decision task. These respond when the current external stimulus matches the one being held in memory, and serve as the decision-making output of the circuit. The frontal region of the model is composed of a 9x9 array of such circuits. Stimuli are encoded by distributed representations of activity in these circuits.

section, and a comparison of different gating mechanisms is performed with the model. Although we do not model the source of the gating, we do make explicit that it operates on the D2 units of the working memory circuit.

Typical electrical behaviors of the memory circuit neurons, as seen in single-cell recordings from monkeys (Funahashi et al., 1990; Goldman-Rakic, 1995), are illustrated in figure 5A. Simulations of a DMS task with the model are shown in figure 5 B and C. In figure 5B, gating is turned on and the stimulus is maintained in the memory during the delay interval. When the matching stimulus appears, the response unit shows a brief response to indicate that a match has occurred. Figure 5C depicts the condition in which no gating is applied, simulating a control condition of passive viewing. In this case, the memory is not maintained during the delay period and no response occurs in the response unit. In single-cell recordings, all of these classes of neurons were found to have at least some degree of stimulus selectivity, in a manner similar to neurons found in more posterior regions of the visual pathway (Funahashi et al., 1990; Wilson et al., 1993). The regions of frontal cortex containing these neurons also appear to be organized in a roughly columnar fashion, with both excitatory and inhibitory neurons in close proximity having similar preferences to stimuli (Wilson et al., 1994; Rao et al., 1999). Each module

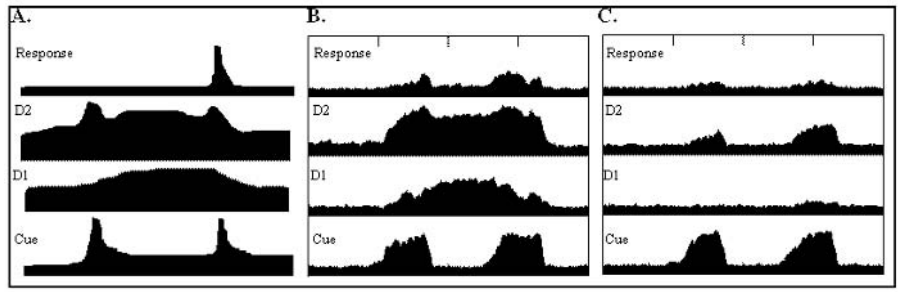


Figure 12.5 Unit electrical activity for each unit type in the frontal cortex over a single trial of the DMS task. A. Data from single-cell recordings in monkeys, in which the monkey was required to maintain a representation of the first stimulus across the delay period (Funahashi et al., 1990; Goldman-Rakic, 1995). B and C. Unit activity in the model working memory circuit with gating turned on (B) and with gating turned off (C), depicting the task condition (B) and control condition (C), respectively. In each figure, the horizontal axis depicts time and the vertical axis depicts the size of the response. Responses to a single instance of the task are shown. A target stimulus is presented first, then there is a delay period followed by the second (test) stimulus. With gating turned ON (5B), the D1 and D2 units maintain high activity during the delay interval. With gating turned OFF, there is no delay period activity. The Cue units respond in each case, but only when the stimuli are present.

of the working memory circuit in the frontal cortex part of the model is composed of four units connected to each other as depicted in figure 4. The frontal area of the model is made up of a 9 x 9 array of such modules, each roughly representing a local assembly such as a column. Representations of objects in the frontal area of the model are formed by distributed activities of these circuits. In addition, a global response unit receives afferents from all individual response units and serves as the decision maker for the model, and is included mainly for assessing the behavior of the model.

Imaging studies of working memory have implicated the frontal cortex as having a major role. In general, frontal regions have higher activation during working memory tasks relative to control tasks that use the same stimuli but do not involve working memory (Haxby et al., 1994; Courtney et al., 1996). One other factor common to many object working memory imaging studies is the enhancement of extrastriate visual areas in the ventral pathway (e.g. Haxby et al., 1994; McIntosh et al., 1996). It is generally thought that both the frontal and posterior associative regions are part of an attention-modulated working memory system.

One unanswered question concerns the mechanism by which modulation of working memory takes place. It is thought that access of items into working memory is mediated by a dopaminergic "gating" signal, although the specific mechanism by which this occurs appears to be complex. Specifically, the D1 dopaminergic receptor subtype appears to enhance the signal-to-noise ratio in neurons of the prefrontal cortex by enhancing stimulus-specific responses and not affecting other response attributes (Desimone and Duncan, 1995). However, too much D1 receptor

action appears to impair working memory, possibly by interfering with the updating process (Williams and Goldman-Rakic, 1995). Note that the naming convention for the D1 and D2 units are taken from Funahashi et al., 1990, and have no relation to the D1 and D2 dopamine receptors. Theoretical models have often used a gating signal, which controls how the contents of working memory are encoded and updated (Zipser, 1991; Cohen and Servan-Schreiber, 1992; Moody et al., 1998; Braver and Cohen, 1999; Braver et al., 1999). In the model of Braver et al. (1999), the gating mechanism was further elaborated by explicitly modeling it as dopaminergic modulation of the frontal cortex. This was implemented as a temporary enhancement of both the afferent excitatory connections into the memory module and the inhibitory connections within the module. This type of complementary/antagonistic mechanism improves the signal-to-noise ratio of responses to attended stimuli, and thus allows access of these stimuli into working memory, and is consistent with the neurobiological evidence.

In our original model (Tagamets and Horwitz, 1998), gating was implemented as a diffuse low-level afferent signal into all the **D2** units in the working memory circuit. Although we did not specify the source of the signal, we did make explicit that it operated on the **D2** type of units only. Recently we implemented a modified mechanism, in which gating is modeled by changes to the dynamics of the units' response properties (Tagamets and Horwitz, 2000). Specifically, gating is effected by increasing the gain and lowering the threshold of the **D2** units of the working memory circuit. The increased gain enhances signal-to-noise ratio by making the responses more robust to above-threshold stimuli and more resistant to subthreshold stimuli, thus making the maintenance of the current item more robust. Lowering the threshold when a new item is about to be encoded increases the access of the item into memory. This mechanism is consistent with the action of dopamine, which has the effect of modulating neuronal responses, rather tham providing additional excitation or inhibition. The different types of units in our model make up a functional circuit in which each unit type has a specific role. The **D2** units serve the dual roles of acting as the link for knowledge of the task by being recipients of the gating, and then maintaining the memory loop (together with the **D1** units) when there is a stimulus to be remembered. The response (**R**) units serve to suppress the memory loop whenever a stimulus is present. They interact competitively with the **D1** units and suppress the **D2** units. The **R** units become fully active only when converging inputs arrive from the external input and the current memory contents, and thus act as the decision-making element. This activity surge also extinguishes activity in the memory loop via inhibitory connections to the **D1** and **D2** units.

There are also feedback interactions of the circuit with other regions. Excitatory projections from the **D2** units into areas IT and V4 enhance the reverberatory circuit that allows stimuli to enter the memory when a stimulus is present. As a result of these connections, some IT neurons tend to also remain active during the initial delay period. Since the **D2** units receive the "voluntary" attentional modulation and serve to integrate the memory with knowledge of the task, it is also possible for self-generated representations to flow backward from the frontal cortex

to visual areas via this route. The **D1** units have diffuse inhibitory back-projections into area IT, which serve to suppress new stimuli when an item is currently being held in memory. This provides some resistance to interference. In a manner similar to single-cell experiments (Miller and Desimone, 1994; Miller et al., 1996), these connections also have the effect of eliminating the delay-period IT activity when a new stimulus occurs.

Simulations were performed by presenting a delayed match-to-sample (DMS) task to the model with and without gating. The latter case serves as a control task. The DMS task was run with two forms of gating: 1) By a diffuse afferent input to all **D2** units, and 2) By increasing the gain and lowering threshold within the **D2** units. The former case, gating via increased afferent activity, models the case in which the gating function is mediated by some other population of cells which signal "attention". In this case, the increased afferent synaptic activity would contribute to a rise in rCBF. In the second case, modulation is achieved by changing the dynamic response properties of the neurons within the memory circuit. This mechanism does not contribute to measured rCBF in itself. Rather, changes in blood flow are due to changes in the synaptic activities within the memory circuit. Two different gating strategies were also tested: 1) Gating was turned on either only during the cue period; or 2) Gating was turned on for the cue period and then maintained for the duration of the trial. The former case represents a gating strategy that only serves to allow updating a new memory, while in the latter case, both updating and maintenance are modulated. The responses illustrated in figure 5B are achieved by the afferent input mechanism applied only during the cue period. Although the memory was maintained, this mechanism did not provide sufficient resistance to interference from intervening stimuli. The goal of the following is to examine alternative gating mechanisms that would provide robustness in a task with intervening stimuli, and concurrently examine the resulting simulated rCBF in order to evaluate which method best approximates human imaging data.

An ABA version of the DMS task was used, in which the appropriate stimulus (A) must be maintained until a matching stimulus occurs, and non-matching stimuli (B) should be ignored. Two criteria were used to evaluate the models: 1) Robustness in the ABA task; this includes the ability to maintain the stimulus across a delay, resistance to intervening stimuli or noise, and appropriate response and clearing of memory when a match occurs; and 2) Quantitative rCBF responses in the frontal area as compared to human imaging data.

12.5.1 Results

Overall, gating by means of gain and threshold modulation over the entire time period showed the best performance. Cues were robustly encoded and maintained in memory until the matching stimulus occurred. The time course from a sample unit is shown in figure 6A. The net effect of this gating mechanism is that responses to optimal stimuli are enhanced while responses to non-optimal stimuli are suppressed, making the circuit less susceptible to noise but at the same time allowing it to

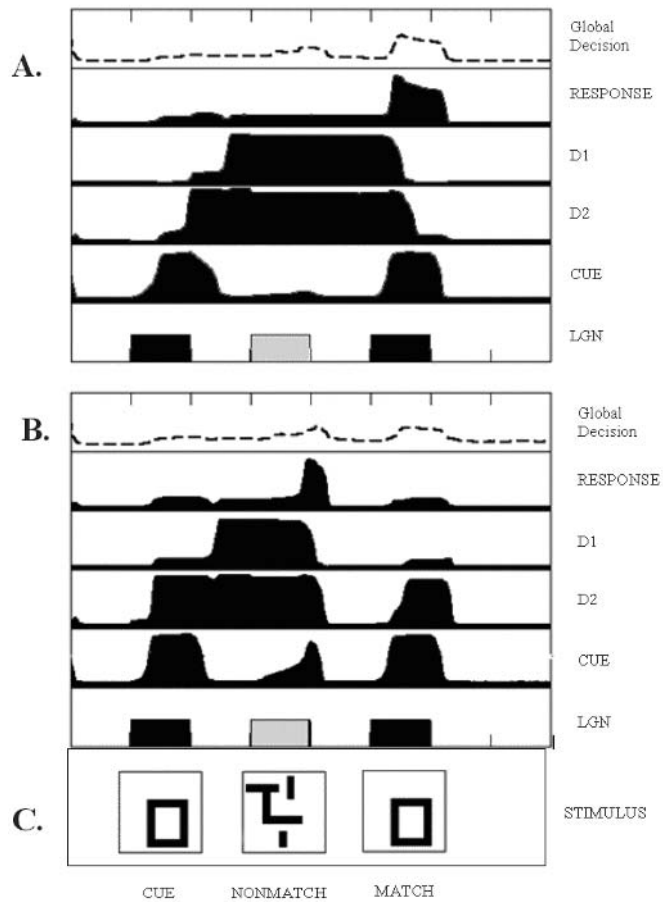


Figure 12.6 Performance of two sample local working memory circuits in the ABA task. A. A unit gated by applying the gain/threshold method for the entire trial period. Here the memory of the original item is maintained across the non-matching intervening stimulus, and an appropriate response is made when the matching stimulus appears. B. A unit from the gating by afferent activation method, again for the entire trial period. All methods produced both types of behaviors, and also cases in which no delay period activity was maintained. In this case, the circuit made an inappropriate decision about the intervening stimulus, and the memory trace was erased. C. Stimuli that were used in the trial. Occurrences are marked in A and B by blocks in the time course labeled LGN. Differences among gating methods were characterized by differences in the total numbers of units which displayed each type of behavior, as indicated by the global decision unit, which sums activities from all response units.

maintain a high level of activity during the delay period. In contrast, the afferent activity model of gating had the poorest performance (figure 6B). This was due mainly to interference from the intervening non-match stimulus, during which many units made inappropriate responses in the R units and recoded the new stimulus into memory. In the other two cases, when gating was active during the initial cue period only, both models had similar moderately good performance, with a slowly

degrading memory activity for many units, but no interference from the intervening stimulus. The similarity makes sense, since during most of the trial period, both had the same parameters after gating was turned off.

The simulated rCBF activity for all four models also differed. Table 12.1 shows the percent increase in Task minus Control subtractions for the four regions of the model, and the results from the data in a face matching study reported by Haxy et al., (1995). The cue-only gain gating had rCBF values that are most similar to human imaging data, in that V4, IT and the frontal region all had similar amounts of increase relative to the control condition, about 6-7%. The whole-period afferent activity gating model deviated most from real data, in that the increase in the frontal area was about 24%, while V4 and IT had increases of 13% and 16%, respectively. Both the afferent cue-only and the gain mediated whole-period rCBF values were similar, with V4, IT and frontal regions having increases of about 9%, 11% and 14%, respectively. While none of the simulated rCBF changes match the experimental data perfectly, there is variability in imaging data as well, as can be seen from the experimental data in Table 12.1. However, it can be seen that the increased gain gating method applied only during the cue period is the closest match to the imaging data. In particular, the other methods produce unrealistically high activation levels in the frontal (PF) region of the model.

Taken together, these results suggest that the gain-mediated model has the most robust behavior and at the same time rCBF values that are most similar to human imaging data. In our implementation, the gating mechanism itself does not contribute directly to the rCBF, since it does not involve afferent synaptic activity. All of the increased rCBF comes from the increases of the units within the memory circuit.

The neuromodulatory actions of dopamine are still poorly understood, but it has become increasingly clear that these are complex and most likely involve other neurotransmitter systems. Computationally, the increased gain and decreased threshold

Table 12.1 Simulated rCBF Results for Different Gating Methods

Simulated rCBF	**P**ercent Change in rCBF			
Gating method	**V**1/V2	**V**4	**I**T	**P**F
Afferent, Cue period only	3%	9%	11%	14%
Afferent, Entire trial	3%	13%	16%	24%
Gain/Threshold Cue only	3%	7%	6%	6%
Gain/Threshold Entire trial	3%	9%	11%	14%
Experimental PET Data				
Face matching – Control (Haxby et al., 1995)	3%	8%	4%	4%
Face matching – Control (Haxby et al., 1994)	2.5%	8%	5%	2%
Face –Location matching (Haxby et al., 1994)	4%	5%	4%	3%

implement competing effects, in that the former increases resistance to noise and the latter makes the circuit more susceptible to noise. Biologically, the evidence suggests that dopamine has both synaptic and extrasynaptic effects on the recipient neurons (Smiley et al., 1994). Furthermore, the synaptic effects are post-synaptic, affecting mainly responses to incoming activity. Since neuroimaging results are thought to reflect mainly presynaptic energy requirements of the transmitter reuptake process (Magistretti and Pellerin, 1999), the implementation of neuromodulators as postsynaptic changes in neuronal dynamics seem appropriate for modeling imaging data until more details of the mechanisms are known.

12.6 Summary and Discussion

Large-scale neural modeling can make explicit use of both anatomical and single-cell data to clarify how the results of these methods can be interpreted in terms of underlying neuronal events. With a large-scale model, explicit hypotheses can be examined in different contexts by varying the effects of parameters that correspond to biological substrates such as synaptic density or receptor efficacy. Manipulating these parameters in a way that changes simulated firing rates to correspond to observed single-cell recordings, then observing the effect on simulated blood flow, may yield insights and help guide experimental design of human studies. For example, inhibition has been hypothesized to raise rCBF and BOLD, making it unclear how to interpret either local increases or the polarity of correlations. The results of the simulation study in section 12.4 suggest that this can be a complex matter that depends on an interaction between local connectivity and task. For example, if the local connectivity is reduced in an area, as it may be in some degenerative disorders, our modeling results indicate that tonic inhibition can produce increases in blood flow under a range of activity levels of the area. With normal local connectivity, however, inhibition may cause decreased rCBF if this area is engaged by a task, and increased rCBF if the task results in little afferent input to the region.

In the second example, a working memory task was used to examine how different types of gating mechanisms might affect both behavior and imaging results. The gain parameter has the effect of increasing signal-to-noise ratio of the simulated electrical activity, which is similar to the proposed effect of dopamine. At the same time, increasing the gain reduces the simulated rCBF in the region. Such manipulations can be used to examine both normal function and intervention studies, such as pharmacological intervention, with applicability to a variety of cognitive disorders.

Other studies with the model have also helped clarify the underlying neuronal mechanisms that generate neuroimaging data. The model has recently been applied to simulate transcranial magnetic stimulation (TMS) data, in which a strong, changing magnetic field applied to a region on the scalp induces intracranial electrical currents that can alter regional neuronal function. TMS has been used in

conjunction with positron emission tomography (PET) to examine inter-regional connectivity of human cerebral cortex (e.g., Paus et al., 1998). We simulated (Nandipati et al., 2000) the effect of TMS using our large-scale neural model, replicating the effects found in Paus et al. (1998), but only when TMS affected primarily the inhibitory units in PFC. In another study, we (Horwitz et al., 1999) used the model to determine what neurobiologic parameters affect the correlations (functional connectivity) of simulated PET-blood flow in different regions. The analysis showed that for two anatomically linked brain regions, the strength of the functional connectivity between them depends both on the strength of their anatomical connectivity (as embodied in the synaptic weights), and also on the extent to which the circuit in which the two regions are nodes is being utilized. All of these studies have helped to clarify the relationship between experimental manipulations and analysis methods, an important factor when interpreting experimental data that is still relatively poorly understood in human imaging.

Acknowledgment

This work was supported in part by the William K. Warren Foundation.

References

R.F. Ackermann, D.M. Finch, T.L. Babb, and J. Engel, Jr. Increased glucose metabolism during long-duration recurrent inhibition of hippocampal pyramidal cells. *Journal of Neuroscience* 4:251-264 (1984).

A. Aersten, G.L. Gerstein, M. Habib, and G. Palm. Dynamics of neuronal firing correlation: Modulation of 'effective connectivity'. *Journal of Neurophysiology* 61:900-917 (1989).

M.A. Arbib, A. Bischoff, A.H. Fagg, and S.T. Grafton. Synthetic PET: Analyzing large-scale properties of neural networks. *Human Brain Mapping* 2:225-233 (1995).

A.L. Bokde, M.A. Tagamets, R.B. Friedman, and B. Horwitz. Functional interactions of the inferior frontal cortex during the processing of words and word-like stimuli. *Neuron* 30:609-617 (2001).

T.S. Braver, D.M. Barch, and J.D. Cohen. Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. *Biological Psychiatry* 46:312-328 (1999).

T.S. Braver and J.D. Cohen. Dopamine, cognitive control, and schizophrenia: the gating model. *Progress in Brain Research* 121:327-349 (1999).

J.D. Cohen and D. Servan-Schreiber. Context, cortex and dopamine: A connectiontist approach to behavior and biology in schizophrenia. *Psychological Review* 99:45-77 (1992).

S.M. Courtney, L.G. Ungerleider, K. Keil, and J.V. Haxby. Relative contributions of occipital, temporal, and frontal cortex to visual perception and working memory: An fMRI study. *NeuroImage* 3:S587 (1996).

R. Desimone and J. Duncan. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience* 18:193-222 (1995).

R.J. Douglas, C. Koch, M. Mahowald, K.A. Martin, and H.H. Suarez. Recurrent excitation in neocortical circuits. *Science* 269:981-985 (1995).

E.N. Eskandar, L.M. Optican, and B.J. Richmond. Role of inferior temporal neurons in visual memory. II. Multiplying temporal waveforms related to vision and memory. *Journal of Neurophysiology* 68:1296-1306 (1992).

K.J. Friston. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping* 2:56-78 (1994).

K.J. Friston. Another neural code? *NeuroImage* 5:213-220 (1997).

K.J. Friston, C. Buechel, G.R. Fink, J. Morris, E. Rolls, and R.J. Dolan. Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* 6:218-229 (1997).

S. Funahashi, C. Bruce, and P.S. Goldman-Rakic. Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. *Journal of Neurophysiology* 63:814-831 (1990).

J.M. Fuster. Inferotemporal units in selective visual attention and short-term memory. *Journal of Neurophysiology* 64:681-697 (1990).

J.M. Fuster, R.H. Bauer, and J.P. Jervey. Cellular discharge in the dorsolateral prefrontal cortex of the monkey in cognitive tasks. *Experimental Neurology* 77:679-694 (1982).

P.S. Goldman-Rakic. Cellular basis of working memory. *Neuron* 14:477-485 (1995).

C.M. Gray, P. Konig, A.K. Engel, and W. Singer. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338:334-337 (1989).

P.E. Haenny, J.H. Maunsell, and P.H. Schiller. State dependent activity in monkey visual cortex. II. Retinal and extraretinal factors in V4. *Experimental Brain Research* 69:245-259 (1988).

J.V. Haxby, C.L. Grady, B. Horwitz, L.G. Ungerleider, M. Mishkin, R.E. Carson, P. Herscovitch, M.B. Schapiro, and S.I. Rapoport. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences - USA* 88:1621-1625 (1991).

J.V. Haxby, B. Horwitz, L.G. Ungerleider, J.M. Maisog, P. Pietrini, and C.L. Grady. The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *Journal of Neuroscience* 14:6336-6353 (1994).

B. Horwitz. Data analysis paradigms for metabolic-flow data: Combining neural modeling and functional neuroimaging. *Human Brain Mapping* 2:112-122 (1994).

B. Horwitz, C.L. Grady, J.V. Haxby, L.G. Ungerleider, M.B. Schapiro, M. Mishkin, and S.I. Rapoport. Functional associations among human posterior extrastriate brain regions during object and spatial vision. *Journal of Cognitive Neuroscience* 4:311-322 (1992).

B. Horwitz and O. Sporns. Neural modeling and functional neuroimaging. *Human Brain Mapping* 1:269-283 (1994).

B. Horwitz, T.W. Long, and M.A. Tagamets. The neurobiological substrate of PET-fMRI functional connectivity. *NeuroImage* 9:S392 (1999).

M. Jueptner and C. Weiller. Review: Does measurement of regional cerebral blood flow reflect synaptic activity?–Implications for PET and fMRI. *NeuroImage* 2:148-156 (1995).

P.J. Magistretti and L. Pellerin. Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. *Philos.Trans.R.Soc.Lond B Biol.Sci.* 354:1155-1163 (1999).

C. Mathiesen, K. Caesar, N. Akgoren, and M. Lauritzen. Modification of activity-dependent increases of cerebral blood flow by excitatory synaptic activity and spikes in rat cerebellar cortex. *J Physiol* 512 (Pt 2):555-566 (1998).

A.R. McIntosh, C.L. Grady, J.V. Haxby, L.G. Ungerleider, and B. Horwitz. Changes in limbic and prefrontal functional interactions in a working memory task for faces. *Cerebral Cortex* 6:571-584 (1996).

E.K. Miller and R. Desimone. Parallel neuronal mechanisms for short-term memory. *Science* 263:520-522 (1994).

E.K. Miller, C.A. Erickson, and R. Desimone. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience* 16:5154-5167 (1996).

S.L. Moody, S.P. Wise, G. di Pellegrino, and D. Zipser. A model that accounts for activity in primate frontal cortex during a delayed matching-to-sample task. *Journal of Neuroscience* 18:399-410 (1998).

G. Nandipati, F.T. Husain, A.R. Braun, M.A. Tagamets, and B. Horwitz. Simulating TMS During PET Using a Large-Scale Neural Network of the Visual System. *NeuroImage* 11:S492 (2000).

T. Paus, R. Jech, C.J. Thompson, R. Comeau, T. Peters, and A.C. Evans. Dose-dependent reduction of cerebral blood flow during rapid-rate transcranial magnetic stimulation of the human sensorimotor cortex. *Journal of Neurophysiology* 79:1102-1107 (1998).

S.G. Rao, G.V. Williams, and P.S. Goldman-Rakic. Isodirectional tuning of adjacent interneurons and pyramidal cells during working memory: evidence for microcolumnar organization in PFC. *Journal of Neurophysiology* 81:1903-1916 (1999).

J.W. Scannell and M.P. Young. Neuronal population activity and functional imaging. *Proc.R.Soc.Lond B Biol.Sci.* 266:875-881 (1999).

J.F. Smiley, A.I. Levey, B.J. Ciliax, and P.S. Goldman-Rakic. D1 dopamine receptor immunoreactivity in human and monkey cerebral cortex: predominant and extrasynaptic localization in dendritic spines. *Proceedings of the National Academy of Sciences - USA* 91:5720-5724 (1994).

L. Sokoloff. Sites and mechanisms of function-related changes in energy metabolism in the nervous system. *Developmental Neuroscience* 15:194-206 (1993).

M.A.Tagamets and B.Horwitz. Modeling brain imaging data with neuronal assembly dynamics, In: Computational Neuroscience: Trends in Research 1997, J.M. Bower (ed). Plenum Press, New York (1997).

M.A. Tagamets and B. Horwitz. Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. *Cerebral Cortex* 8:310-320 (1998).

M.A. Tagamets and B. Horwitz. A model of working memory: Bridging the gap between electrophysiology and human brain imaging. *Neural Networks* 13:941-952 (2000).

J.G. Taylor, B. Krause, N.J. Shah, B. Horwitz, and H.W. Mueller-Gaertner. On the relation between brain images and brain neural networks. *Hum.Brain Mapp.* 9:165-182 (2000).

L.G.Ungerleider and M.Mishkin. Two cortical visual systems, In: Analysis of Visual Behavior, D.J. Ingle et al. (eds). MIT Press, Cambridge, MA (1982).

G.V. Williams and P.S. Goldman-Rakic. Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature* 376:572-575 (1995).

F.A. Wilson, S.P. Scalaidhe, and P.S. Goldman-Rakic. Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* 260:1955-1958 (1993).

F.A.W. Wilson, S.P. Scalaidhe, and P.S. Goldman-Rakic. Functional synergism between putative gamma-aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex. *Proceedings of the National Academy of Sciences - USA* 91:4009-4013 (1994).

D. Zipser. Recurrent network model of the neural mechanism of short-term active memory. *Neural Computation* 3:179-193 (1991).

This page intentionally left blank

13 Large-Scale Networks in Learning Analyzed with Partial Least Squares

Hemanth P. Nair
Center for Behavioral Neuroscience
Emory University
Atlanta, Georgia

Francisco Gonzalez-Lima
Institute for Neuroscience and Department of Psychology
University of Texas at Austin
Austin, Texas

The development of the ability to express learned behavior during the postnatal period is presumably related to maturational changes in the recruitment of particular neural systems to guide the behavior. By assessing brain functional activity during transitional periods of behavioral development, we may gain valuable insight into when particular neural systems come on-line and impact behavior. These issues were investigated by applying Partial Least Squares (PLS) analysis to metabolic mapping data obtained from developing rats. Preweanling rat pups aged postnatal day 12 (P12) and P17 were trained on two different instrumental reward schedules, injected with fluorodeoxyglucose (FDG), and then shifted to continuous nonreward (extinction). Behavior during extinction varied with training in P17 pups but not P12 pups. In the first application of PLS, an analysis analogous to a traditional univariate means analysis was performed to identify large scale networks, within 39 regions of interest, either commonly activated across groups or which differentiated groups. A second application of PLS was used to identify dominant patterns of covariances between regions (i.e. functional connectivity) that distinguished training and age groups.

13.1 Introduction

Electrophysiological and imaging studies have supported the view that learning and memory proceeds through the concerted activity of distributed neural systems. Ex-vivo metabolic mapping and in-vivo functional imaging techniques aid the identification of such systems by virtue of their ability to measure, simultaneously, the activity of all parts of the brain during a particular task. However, efficiently extracting information about those systems from such a large data set - particularly how they relate to experimental condition or a behavioral measure poses a considerable statistical challenge. Here we demonstrate the utility of a multivariate technique known as partial least squares (PLS) for this purpose. The PLS method has its roots in a family of least squares models of correlation matrices introduced by Sewel Wright in the 1920s. The specific computation described here, however, was fully developed by Bookstein and colleagues (Bookstein, 1994) in studies of alcohol teratogenesis. McIntosh et al. (1996) were the first to adapt the technique to human functional imaging data; in this case PLS identifies spatial patterns of functional activation (singular images) that covary with task, behavior, or other regions of interest (ROIs). In this chapter, we demonstrate how the PLS method may be used for the analysis of brain metabolic mapping data of learning and postnatal development.

13.2 Conceptualization of PLS

At the core of the PLS method is the singular value decomposition (SVD) of the cross-correlation matrix derived from two blocks of variables. The cross-correlation matrix is part of a correlation matrix in that it includes only correlations between the two blocks of variables rather than variables within a block and is usually not square (hence, SVD or least squares analysis of part of a correlation matrix leads to the name partial least squares Bookstein, 1994). The output is a series of latent variables (LV) describing the covariance between linear combinations of the original variables. Each LV accounts for successively less of the original cross block covariation. Thus, PLS identifies a new set of variables that optimally relate the variables in the original two blocks using the fewest dimensions (McIntosh et al., 1996). For example, suppose brain functional activity is measured across three conditions or tasks. Two possible design effects can be represented by a set of orthogonal vectors (2 1 1, 0 1 1). These vectors constitute a block of variables (call them A) that are correlated with another set of variables (B) the functional activation measured in each region of interest (ROI). SVD of this cross correlation matrix computed from A and B will yield a new linear combination of the design variables that optimally covary with a new linear combination of ROIs. The output thus identifies sets of regions associated with a particular design effect; basically, sets of regions showing particular mean differences (or commonalities, which will

be illustrated later) and the nature of those differences. Hence PLS applied in this way, also known as design PLS, can be conceptualized as a multivariate extension of a traditional means analysis (analysis of functional activation).

Suppose the design variables in A in the previous example are replaced by a separate set of ROIs. In this case, the same computation results in a multivariate extension of covariance analysis (analysis of functional connectivity see Horwitz et al., 1992 ; McIntosh and Gonzalez-Lima, 1998; Nair and Gonzalez-Lima, 1999). This seed PLS will identify those regions in B with which set A demonstrates task related covariance changes and the nature of those differences (the particular group differences and the direction of the changes) (McIntosh and Gonzalez-Lima, 1998; Nair et al., 2001a). Alternatively, if A is now a behavioral measure the seed PLS identifies sets of regions optimally covarying with performance and the particular nature of the group similarities or differences in that relationship. In summary, the power of PLS as applied to imaging data resides in its ability to identify, usually in a single omnibus step, sets of regions showing common activational effects, interregional covariance, or covariation with performance. The following sections will detail the PLS method through specific examples derived from our studies of learning and postnatal development. The goal is to present step by step computation of the PLS analysis. For a more thorough look at the mathematical basis of PLS please refer to Bookstein (1994) and McIntosh et al. (1996).

13.3 PLS in Studies of Functional Development of the Brain and Behavior

If distributed neural systems guide associative learning in adults then the presence or absence of the learning during the postnatal period presumably relates to the maturational state of those systems. Thus, as a behavior emerges over the course of the postnatal period, it may be inferred that the neural systems and, importantly, the functional interactions between neural systems that support the behavior are maturing. By combining brain functional imaging with behavior, our aim has been to gain insight into developmental changes in the neural systems that are on-line and operable at successive ages of development. In preweanling rats, the ability to readily modify behavior in response to different reinforcement schedules changes with development (Lilliquist et al., 1999). Postnatal day 17 (P17) rats trained in a straight alley runway on an alternating schedule of reward and nonreward, (patterned single alternation or PSA) rapidly decrease their responding when switched to continuous nonreward (extinction). If animals are trained on a pseudorandom reward schedule (pseudorandom partial reinforcement or PRF), they are slow to inhibit responding (i.e. they are persistent) during extinction relative to P17 PSA-trained pups. Thus their prior learning during acquisition guides P17 extinction behavior. In younger, P12 pups, PSA and PRF trained subjects display essentially the same acquisition behavior as P17 pups. However, they show no differences in responding during extinction. Thus, the younger pups ability to utilize

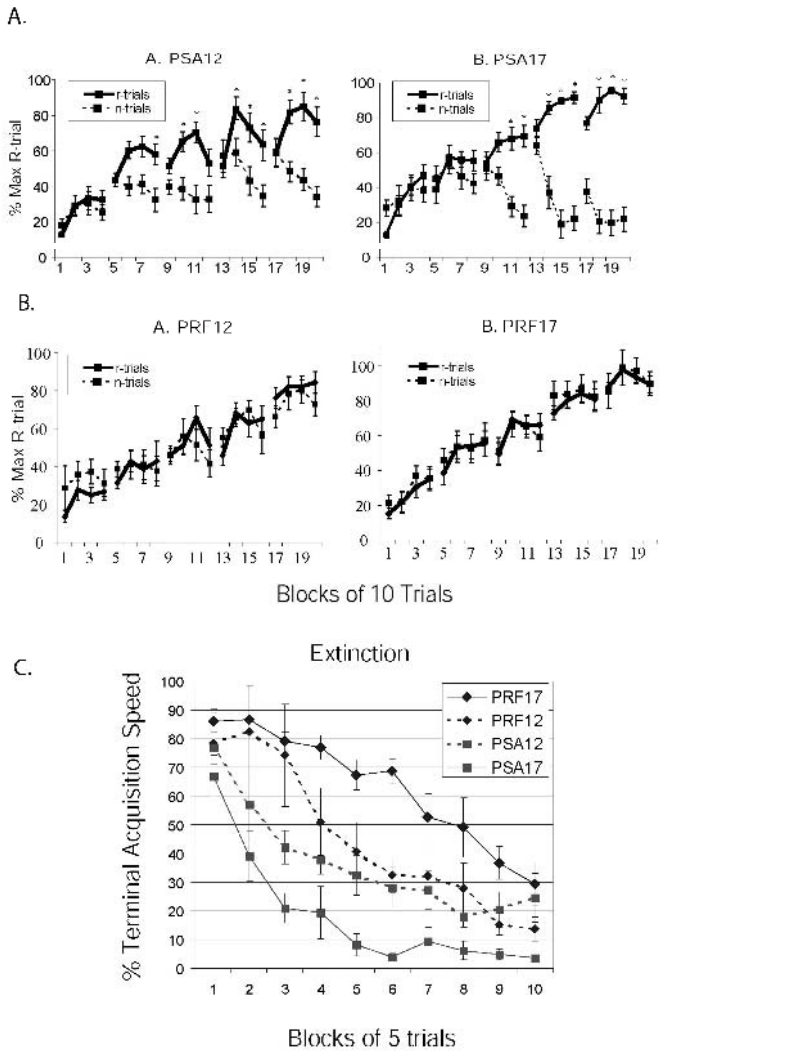


Figure 13.1 Behavioral Effects. Please refer to Nair et al. (2001b) for full details on the analysis of behavioral effects. Behavioral training lasted two days. 200 trials of acquisition (five sessions of 40 trials) followed by 50 trials of extinction were conducted. A) PSA Acquisition Effects. Both PSA12 and PSA17 pups discriminated reward and nonreward trials by the end of training. The breaks in the data indicate the five sessions of acquisition. * indicates significantly increased run speeds on reward trials relative to nonreward trials ($p < 0.05$, using repeated measures ANOVA). B) PRF Acquisition Effects. PRF12 and PRF17 demonstrated no differences in run speed between rewarded and nonreward trials. C) Extinction Effects. PSA17 pups demonstrated significantly faster extinction rates relative to PRF17 pups, based on a nonlinear regression analysis of run speeds (Nair et al., 2001b). No differences were found in the P12 groups. FDG was administered at the start of the extinction session, which lasted 50 minutes. A group of handled control animals (N=10) was included at each age to control for handling effects.

learning established during acquisition appears to be compromised when shifted to extinction. These behavioral effects are summarized in Figure 13.1.

These data are important because they suggest a maturational shift has occurred in the neural systems engaged during extinction training. Functional systems related to flexibility in behavior might be more mature in P17 pups compared to P12 pups. Moreover, it is possible that P12 pups are relying on a fundamentally different neural system. Alternatively, there could be common systems engaged at both ages, but the recruitment of those related to flexible responding distinguish the older pups. To address these issues, we used fluorodeoxyglucose (FDG) autoradiography to map metabolic activity during extinction in the two age groups. FDG, a radiolabeled glucose analog, can be used to measure regional changes in brain metabolic activity occurring during behavior. Since energy utilization and functional activity are closely correlated (Sokoloff, 1992), FDG serves as an index of brain functional activity (see Gonzalez-Lima (1992) for review of FDG methodology and applications). A design PLS was used to identify regions showing age-related and task related changes in activation. Seed PLS was then used to identify regions showing significant covariance changes with three frontal cortical regions that were hypothesized to be important for flexible responding during extinction.

13.4 Design PLS: Activational Changes due to Age and Learning

A univariate covariance analysis previously showed that functional interactions between the septo-hippocampal limbic system and other regions, such as the ventral tegmental area, are quite different in P17 PSA, relative to P12 PSA pups during extinction (Nair and Gonzalez-Lima, 1999). While functional coupling among these regions occurred in P17 PSA pups, they were functionally dissociated in P12 PSA pups. The functional dissociation of the septo-hippocampal system in P12 pups implies that a different neural system is guiding their behavior. General motor coordination and somatosensation are among the first functions to appear in the behavioral ontogeny of the rat (Almli and Fisher, 1977) and so it was hypothesized that P12 pups may be relying on somatosensory and motor regions. Furthermore, studies assessing developmental changes in baseline metabolic activity indicate that regions more caudal on the neuroaxis attain functional maturity earlier than more rostral areas, particularly forebrain structures (Chugani et al., 1991; Kennedy et al., 1972; Nair et al., 1999). Hence, somatosensory, motor, and brainstem regions may form the dominating circuit in P12 behavior while P17 pups may have shifted to relying on more rostral brain regions, such as frontal and limbic structures. Alternatively, the same brainstem/sensory-motor system may be operating in both P12 and P17 pups, but the additional recruitment of frontal and limbic structures may allow for the differential responding in P17 pups. Hence, there may be common regions operating in both P12 and P17 pups, but the additional frontal and limbic recruitment in the P17 pups allows for their differential responding. The design PLS analysis of activational effects is uniquely suited to answer these questions, as it

identifies sets of regions which demonstrate both group commonalities or differences in functional activation, in one omnibus step (McIntosh et al., 1996). The capability to identify systems commonly activated across age or task thus distinguishes the PLS method from a univariate means analysis, which emphasizes differences between groups. Furthermore, despite the various hypotheses proposed above, conducting a univariate analysis in a hypothesis driven manner is less powerful than the data-driven approach offered by PLS. In the univariate case, limiting the analysis to hypothesized regions, though preserving statistical robustness, may result in excluding regional effects that might be present in the data. Alternatively, one could perform the univariate analysis on a large number of regions. However, the propensity for a Type I error would be greatly inflated on the one hand, but using a correction would be overly conservative on the other. By virtue of its treatment of regions simultaneously and its ability to identify main effects and interactions in one omnibus step, the data-driven, PLS approach is better suited for this kind of data set.

13.4.1 Methods

13.4.1.1 Behavioral Training

Animals were given acquisition training (as described in Figure 13.1), injected with FDG and then extinguished. FDG incorporated during extinction in P12 and P17 PSA-, PRF-trained and handled control (HC) pups was measured via quantitative image analysis in 39 regions throughout the rostral-caudal extent of the brain. HC pups were exposed to the same environmental conditions and given the same amount of reward but outside the runway apparatus. They were merely placed in and out of the runway for acquisition and extinction sessions.

13.4.1.2 FDG Method

Immediately prior to the extinction session subjects were injected intraperitoneally with $18Ci/100g$ b.w. of $[^{14}C(U)]$ fluoro-2-deoxy-glucose (FDG); (specific activity=300 mCi/mmol, American Radiolabeled Chemicals) in 0.1 ml of physiological saline. Animals were trained for approximately 50 min., the duration of the extinction session. Upon completion of the test period, the animal was removed from the chamber and rapidly decapitated. The brain was then quickly removed and frozen. Sections of the brain at $40\mu m$ were taken in a cryostat at $-20^{o}C$. The FDG slides, along with plastic micro-scale standards of known 14C concentrations (Amersham) were exposed to Kodak EB-1 film for 2 weeks and then developed. Images from the film were placed on a DC-powered light box and captured through a black-and-white video camera (Javelin JE2362). Incorporation of FDG was quantified using JAVA image analysis software (version 1.4, Jandel Scientific Corp.). A calibration curve was created based on the absolute gray levels of the 14C standards on the film. Subsequent densitometric measures taken from brain images were then auto-

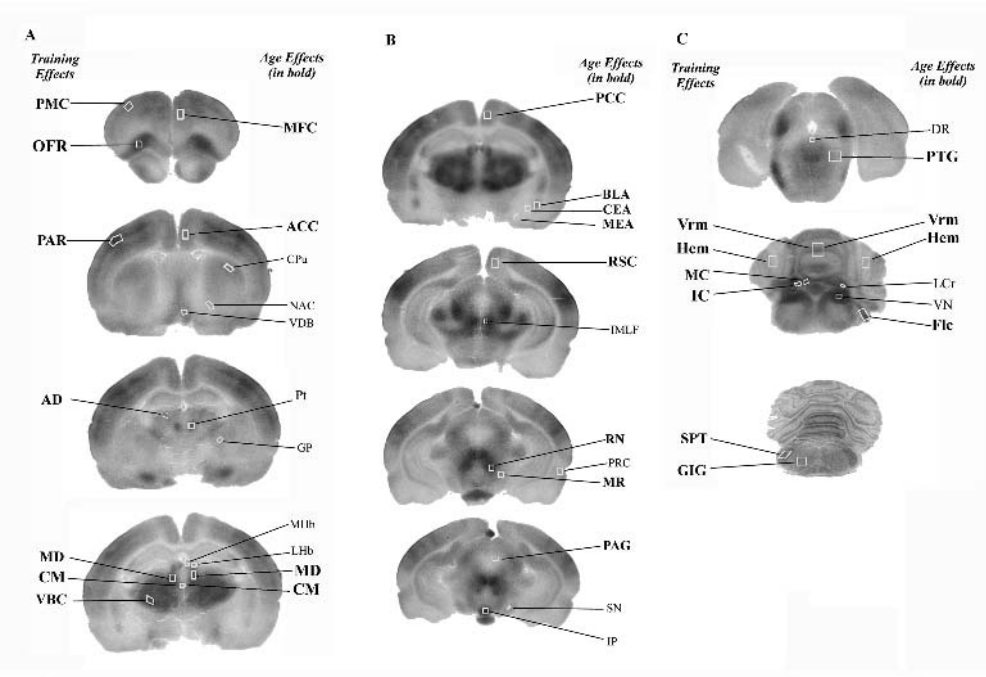


Figure 13.2 FDG autoradiographic images of P17 brain sections depicting regions of interest and effects revealed by PLS. Training effects are indicated on the left side of the images while age related effects are indicated on the right. Regions significantly associated with the effects are indicated in bold. Those that were sampled but demonstrated no effect are in plain text, on the right

matically expressed in terms of isotope incorporation per gram of tissue (nCi/g).

Details of the behavioral effects and design, FDG administration, and image analysis have been reported (Nair et al., 2001b). Approximate sampling levels and anatomical location of readings are illustrated in Figure 13.2A, 13.2B and 13.2C.

13.4.1.3 Design PLS

A preliminary analysis indicated that a single analysis on the entire data set was problematic even for PLS, due to the large number of groups. Therefore the ROIs were regrouped according to anatomical/functional characteristics and separate PLS analyses were performed on each group. The 5 sets of regions were grouped as limbic thalamic/cortical regions, somatosensory/motor regions, basal forebrain, brainstem, and cerebellum (see Table 13.1). The details of the PLS analysis follows and are illustrated in Figure 13.3.

Step 1. Design-ROI Cross Product Matrix. A cross product matrix (S) between the matrix containing the FDG uptake values, in which columns represent the ROIs and rows represent subjects, and a matrix containing design contrasts

(orthogonal Helmert contrasts) was calculated. The design contrasts are dummy variables coding for the design features. For example, given three groups of three subjects each, the contrast (2 2 2 -1 -1 -1 -1 -1 -1) distinguishes group 1 from 2 and 3 while (0 0 0 -1 -1 -1 1 1 1) contrasts groups two and three (Figure 13.3A). In this case, six contrasts were used. The first contrast distinguished group 1 from the other 5 (i.e. 5 -1 -1 -1 -1 -1). The next contrasted group 2 from the remaining four (4 -1 -1 -1 -1), and so on.

Table 13.1 FDG Uptake Means ± Standard Error (nCi/g)

Limbic Thalamic /Cortical	**ALT12**	**PRF12**	**HC12**	**ALT17**	**PRF17**	**HC17**
Orbital Frontal Cortex (OFC)[b]	1311±96	1360±94	1077±10	31518±110	1429±85	1114±126
MedialFrontal Cortex (MFC)[a]	562±56	554±45	478±51	1107±99	1055± 56	855±98
Anterior Cingulate Cortex (ACC)[a]	686±61	672±60	516±70	1025±93	1059±68	835±108
Posterior Cingulate Cortex (PCC)[a]	639±55	673±55	558±68	1113±104	1026±71	876±106
Retrosplenial Cortex (RSC)	483±46	501±35	401±50	869±78	755±51	695±82
Perirhinal Cortex (PRC)[a]	389±36	384±26	310±35	550±40	530±36	503±52
Medial Dorsal Nucleus (MD)[a,b]	1024±76	1146±85	924±102	1561±103	1427±92	1138±122
Centromedian Nucleus (CM)[a,b]	639±51	658±57	518±56	1044±69	1009±63	801±88
Anterior Dorsal Nucleus (AD)[c]	1038±96	748±63	762±55	954±87	861±72	843±116
Paratenial Nucleus (Pt)	894±115	669±66	698±76	893±113	896±126	772±102
Somato sensory/Motor						
Primary Motor Cortex (PMC)	1036±100	972±88	691±54	982±77	1006±72	888±117
Anterior Parietal Cortex (Par)[b]	1454±94	1399±98	1079±82	1308±96	1325±91	1097±126
Globus Pallidus (GP)	654±40	748±52	613±48	729±94	731±44	658±67
Ventrobasal Complex (VBC)[b]	1471±60	1635±91	1235±104	1490±86	1460±80	1136±130
Red Nucleus (RN)	938±78	972±72	768±85	1203±82	1243±6	9968±105
Substantia Nigra (SN)	469±36	508±37	396±38	692±57	650±52	565±63
Spinal Trigeminal Nucleus(SpT)[b]	1151±69	1229±85	760±92	1065±77	1061±79	860±110
Vestibular Nucleus (VN)[b]	1533±82	1640±90	1215±85	1556±88	1545±78	1235±142
Caudate Putamen (CPu)	848±91	924±99	655±82	1088±121	1035±76	787±94
Basal Forebrain						
Ventral Diagonal Band (VDB)	775±74	830±87	683±65	1062±99	987±68	792±102
Basolateral Amygdala (BlA)[a]	456±32	507±38	424±48	819±72	759±45	682±83
Central Amygdala (CeA)[a]	350±23	385±31	301±34	500±31	504±27	460±47
Medial Amygdala (MeA)[a]	386±36	420±39	334±34	537±49	561±38	501±55
Nucleus Accumbens (NAC)	478±36	627±57	496±47	644±85	645±45	581±53
Medial Habenula (MHb)	846±59	914±70	752±77	1088±87	1040±64	865±85
Lateral Habenula(LHb)	863±69	941±66	768±77	1185±86	1125±60	962±110
Brainstem						
Periaqueductal Grey (PAG)[a]	387±33	404±30	343±48	582±38	558±37	531±53
Mesencephalic Reticular Nucleus (MR)[a]	581±45	613±53	462±55	875±68	860±56	756±77
Interpeduncular Nuclues (IP)	1120±75	1200±80	1007±97	1343±10	21282±117	1096±118
Interstitial Nucleus, Medial Longitudinal Fasciculus (IMLF)	1035±85	1029±76	826±76	1439±105	1389±91	1118±121
Dorsal Raphe (DR)	466±44	507±40	408±47	639±40	606±43	561±57
Pedunculopontin Tegmental N. (PTG)[a]	565±51	580±45	427±48	795±59	770±54	697±83
Gigantocellular nucleus (GIG)	791±62	847±59	594±54	892±63	867±58	724±90
Cerebellum						
Medial Cerebellar Nuclues(MC)[a,b,d]	1149±70	1240±69	940±88	1372±99	1357±82	1102±144
Interpositus (IC)[d]	1077±85	1097±75	834±78	1356±110	1330±74	1086±139
Lateral Cerebellar N (LCr)	950±74	990±78	749±78	1318±105	1318±66	1049±136
Cerebellar Vermis (Vrm)[a,b]	839±59	898±56	686±56	1201±87	1178±69	914±102
Cerebellar Hemisphere (Hem)[a,b]	452±41	504±40	342±28	726±67	723±59	561±72
Flocculus (Flc)[a]	550±32	579±60	547±70	945±93	1007±76	754±107

Notes: Label *a* indicates increased activity in all three P17 groups relative to P12 groups; *b* stands for increased uptake in ALT and PRF pups relative to HC, in both age groups; and *c* denotes increased uptake in ALT12 and PRF12 pups relative to HC12 pups.

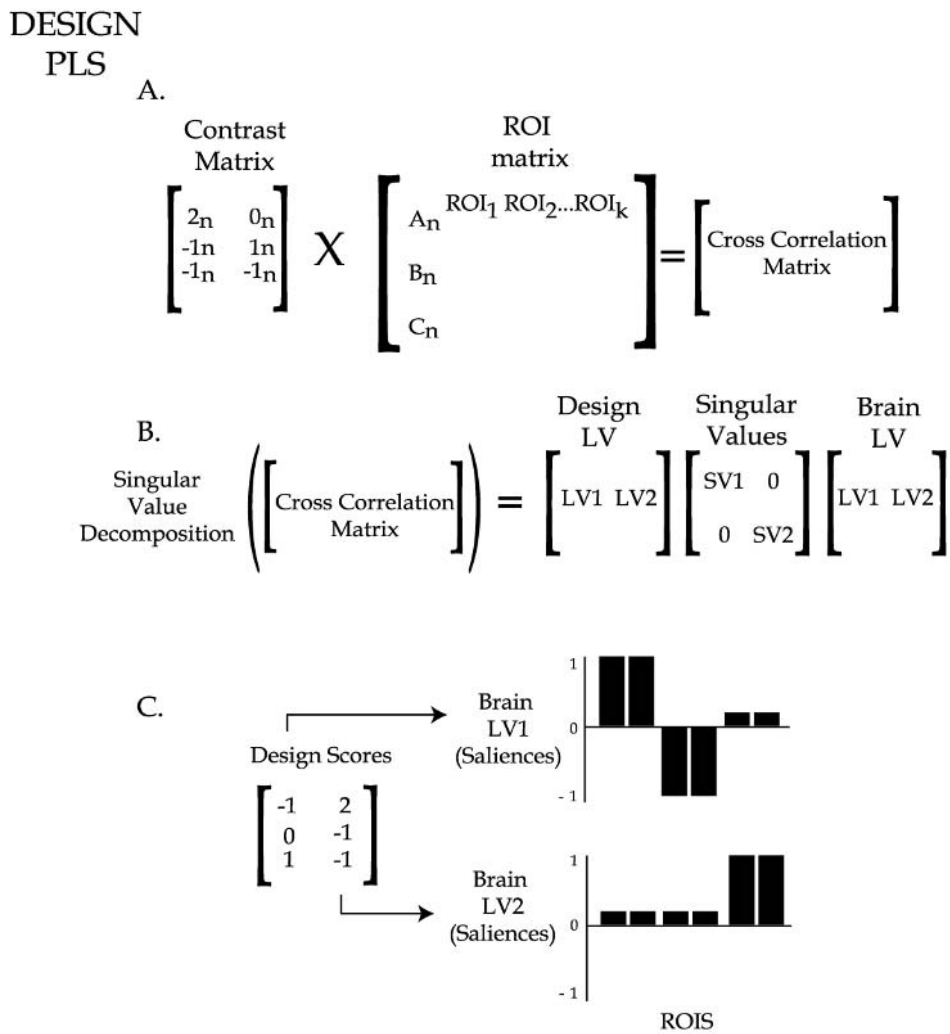


Figure 13.3 A) The design PLS analysis begins by computing a cross correlation matrix between design contrasts (contrast matrix) and the regions of interest (ROI matrix). B) This matrix is then decomposed using singular value decomposition (SVD), which returns mutually orthogonal paired latent variables (LV). Each LV describes a set of weights or saliences associated with the original design contrasts (design LV) and are maximally correlated with a corresponding set of saliences associated with the ROIs (brain LV). The magnitude of the covariance for each LV is indicated by the singular values. C. Suppose 6 ROIs are sampled in three experimental groups of animals and the largest proportion of the cross-block covariation is accounted for by a contrast between group 1 and 3 across regions 1-4. The first latent variable (with the largest singular value) will consist of high saliences for regions 1 through 4; the contrast saliences when multiplied by the original contrasts will result in a new set of contrasts. The first vector in the design score matrix thus corresponds to the first set of saliences (Brain LV1), as indicated by the arrow. Suppose the next greatest proportion is a contrast between groups 1 vs. 2 and 3 on the last two ROIs. When the contrast saliences are multiplied by the original contrasts, the new design scores will approximate the 2 1 1 contrast. The ROI saliences will be high for the last two regions. Thus, the second vector in the design score matrix corresponds to Brain LV2, as indicated by the arrow.

Step 2. SVD of S. SVD summarizes large covariance structures in terms of a smaller number of components (Reyment et al., 1996). The cross product matrix S was decomposed via SVD into a series of latent variables (LV). Each LV describes a set of weights or saliences associated with the original design contrasts and are maximally correlated with a corresponding set of saliences associated with the ROIs. The covariation between the vectors comprising each LV is returned as a third PLS descriptor called the "singular value". The sum of squares of the singular values equals the sum of squares of the original cross block correlation matrix. Hence, each singular value indicates the proportion of the cross block correlation (SSCB) accounted for by the covariation between the region and design saliences. The successive singular values account for the covariances in decreasing order of importance, that is, the largest amount of the SSCB is carried by the first LV, the second largest amount by the third, and so on (Figure 13.3B).

Step 3. Design Scores. By multiplying the contrast saliences by the original design contrasts, we obtained design scores. The design scores are a linear combination of the original design contrasts and indicate a new set of contrasts maximally correlated with the ROI saliences. Thus, those regions showing high saliences are most associated with the new set of contrasts indicated by the design scores (Figure 13.3C).

Step 4. Statistical Significance. The statistical significance of each LV pair can be assessed by bootstrap methods or permutation tests (McIntosh et al., 1996). Here we chose the latter. The subject to group assignment was randomized, and the PLS recomputed. This was repeated 5000 times and the probability of a singular value greater than or equal to the original was computed. Those singular values with a probability of less than 0.05 were considered significant. The individual saliences were tested in the same way. Testing of individual saliences is somewhat controversial, however, as it does not take into consideration the strongly dependent nature of the individual saliences. Therefore, it has been suggested that significance tests be limited to the singular values (Bookstein, personal communication).

13.4.2 Results

The interpretation of the design PLS output involves relating the design scores to the regional saliences. Those saliences that are significant demonstrate significant covariation with the corresponding design scores. The salience-design score relationship essentially describes the nature of the mean differences in FDG activity across groups. Saliences can be either negative or positive; the sign is interpreted in relation to the signs of the values of the design score. A negative region salience and design scores of (3 -3 0.01) describes a pattern of FDG uptake which contrasts groups 1 and 2, such that group 2 demonstrates greater uptake than group 1. A positive salience in the same case indicates lower uptake in group 2 relative to group 1. Design scores, the effect described by the design scores, the regions significantly

Table 13.2 Age-Related Effects: Design PLS

Regions		Design Scores [ALT, PRF, HC]	Effect	Signficant Regions +Salience	P % SSCB
Limbic/Cortical	LV1	[-2.0, -1.5, -1.3](P17) [2.4, 1.2, 1.3](P12)	P17>P12	MFR (-0.50) RS (-0.33) ACG (-0.36) MD (-0.39) PCG (-0.41) Cm(-0.37)	p<0.01 86.9
	LV2	[-.09, .28, .43](P17) [-4.3, 2.1, 1.5](P12)	ALT12 > PRF12, HC12	AD (-0.78)	p< 0.01 86.9
Basal Forebrain	LV1	[-1.9, -1.5, -1.3](P17) [2.8, 0.48, 1.3](P12)	P17 > P12	BLA (-0.56) MEA (-0.23) CEA (-0.27)	p < 0.05 90.1
Brainstem	LV1	[-1.8, -1.4, -1.3](P17) [2.0, 0.78, 1.8](P12)	P17 > P12	PAG (-0.30) IMLF (-0.27) MR (-0.47) PTG (-0.39)	p < 0.01 94.5
Cerebellum	LV1	[-1.8, -1.5, -1.2](P17) [2.2, 0.7, 1.7](P12)	P17 > P12	LC (-0.48) CH (-0.35) CV (-0.43) Flc (-0.48)	p < 0.01 93
	LV2	[.30, 0.89, 1.1](P17) [-2.4, -2.4, 2.5] (P12)	ALT12, PRF12 > HC12	IC (-0.37) MC (-0.58)	p< 0.01 6.5

associated with the effect for each PLS analysis, p values, and percent SSCB, are presented in Table 13.3. While up to 3 significant LVs were obtained in some analyses the third usually accounted for less than 5% of the cross block correlation, usually involved only one region, and either represented theoretically unmeaningful contrasts (e.g. PSA12 vs. HC17), or statistical noise. We thus constrained our significant findings to the first two significant LVs for each analysis.

Limbic Thalamic/Cortical Regions. In the medial frontal, anterior cingulate, posterior cingulate, and retrosplenial cortices and the medial dorsal and centromedian thalamic nuclei, FDG uptake was significantly increased in all P17 groups relative to the P12. The first two LVs were significant and accounted for 95.8% of the sum-squared cross block correlation (SSCB). Thus, the first LV ($p < 0.01$, 86.9% SSCB) identified a pattern of activity differentiating the two age groups.

As indicated in Table 13.2, the design contrasts associated with this effect were [-2.0, -1.5, -1.3] (P17) and [2.4, 1.2, 1.3] (P12). The three numbers in each vector correspond to PSA, PRF and HC groups, respectively. Note that the contrasts are approximately the same magnitude in each case, but opposite in sign. Thus, these contrasts indicate an age related difference that is common to all three experimental groups. The aforementioned regions significantly associated with this contrast demonstrate salience values negative in sign. This means that FDG uptake is higher in the P17 groups relative to the P12 groups. The magnitude of the salience itself is proportionately related to the magnitude of the effect associated with the region. The second LV reflected differences among the P12 groups. LV2 ($p < 0.01$, 8.9% SSCB) described significantly increased FDG in the anterior dorsal thalamic nucleus. Note that in this case, the design scores are much smaller in the P17 group ([-0.09, 0.28, 0.43]) relative to the P12 group ([-4.3, 2.1, 1.5]). This indicates that the changes involving the anterior dorsal thalamic nucleus are associated with the P12 group. The P12 contrast distinguishes the PSA12 pups from the PRF12 and HC12 pups. Because the associated salience for the anterior dorsal nucleus is -0.78, FDG uptake is higher in PSA12 pups relative to the PRF12 and HC12 groups.

Basal Forebrain. LV1 ($p < 0.05$, 90.1% SSCB) identified an age-related contrast in which the amygdalar nuclei (basolateral, central, and medial amygdala) were increased in all three P17 groups compared to P12 groups. For LV1 the contrasts ([-1.9, -1.5, -1.3] (P17), [2.8, 0.48, 1.3] (P12)) are opposite in sign and comparable in magnitude across the age groups, indicating an age effect. On the other hand, note that in the design scores for the P12 group, the value associated with the PRF group (0.48) is relatively small compared to the PSA and HC groups. This is due to the fact that the PRF12 group tended to have higher values across the amygdaloid regions compared to the PSA12 and HC12 group, resulting in less difference from the PRF17 group. Hence the reduced design score for the PRF12 group was due to the fact that it did not show the same magnitude of decrease from its P17 counterpart, as did the PSA12 and HC12 groups. The largest regional salience was associated with the BLA, which indicates that it demonstrated the greatest difference between P12 and P17 pups.

Brainstem. LV1 ($p < 0.01$, 94.5% SSCB) attained significance among these regions and reflected an increase between P12 and P17 groups in the periaquaductal gray, mesencephalic reticular nucleus, and the pedunculopontine nucleus. The contrasts for the P17 ([-1.8, -1.5, -1.2]) and P12 ([2.2, 0.7, 1.7]) were approximately equal in magnitude but opposite in sign, indicating the age difference. Again, the lower design score in the PRF12 case indicates that the magnitude of the age difference was less in the PRF case, due to increased uptake in the PRF12 group relative to the PSA12 and HC12. The saliences associated with the periaquaductal gray (-0.30), mesencephalic reticular nucleus (-0.47), and the pedunculopontine nucleus (-0.39) indicated the direction of the effect (negative saliences so greater uptake in P17 pups).

Cerebellum. LV1 and LV2 were significant for the cerebellar regions and accounted for 99.5LV2 ($p < 0.01$, 6.5%) identified increased FDG in the interpositus and medial cerebellum which distinguished the PSA12 and PRF12 groups from the HC12 group.

13.4.2.1 FDG Changes Using Residualized Values

The PLS methodology identifies the most dominant pattern of changes across experimental conditions. The results from the first PLS analysis revealed a dominant age-related effect, wherein all P17 groups demonstrated higher FDG uptake values relative to P12 groups across several ROIs. It is possible that other effects interactions may not have attained statistical significance due to the dominant age effect. Therefore, we performed a residual analysis in which the age effect was removed from the data, using linear regression, and PLS was performed on the remaining residual values. The aim was to identify interactions that may have occurred independent of the main effect of age. ROI uptake values were first regressed on to the contrast vector [1 -1] (coding for age P12 vs. P17) and a regression equation

Table 13.3 Training Effects: Residual Design PLS

Regions		Design Scores [ALT, PRF, HC]	Effect	Signficant Regions +Salience	P % SSCB
Limbic/Cortical	LV1	[-1.4, -1.0, 2.4](P17) [-0.73,-0.75,1.5](P12)	PREF,ALT >HC	OFC (-0.55) MD (-0.49) CM (-0.32)	p < 0.01 91.6
Samatosensory/ Motor	LV1	[-0.88, -0.85, 1.7](P17) [-1.0,-1.2,2.2](P12)	PREF,ALT >HC	Par (-0.37) SPT (-0.42) VBC (-0.42) VN(-0.45)	p < 0.01 95.4
Brainstem	LV1	[-1.1, -0.9, 2.0](P17) [-0.87,-2.2,1.9](P12)	PREF,ALT >H C	IMLF (-0.61) Gig (-0.46)	p < 0.02 96.5
Cerebellum	LV1	[-1.1, -1.1, 2.2](P17) [-0.76,-0.98,1.7](P12)	PREF,ALT >HC	CV (0.44) MC (-0.48) CH (-0.28)	p < 0.01 96.4

derived (wherein contrast is the predictor variable X). The regression equation was applied to the contrast vector to derive a predicted set of ROI values. The residual values were then obtained by subtracting the predicted from original ROI values.

In all sets of regions except for the basal forebrain, the first LV was significant and indicated an increase in FDG uptake among PRF and PSA pups relative to HC pups at both ages. The same regions showed this effect in both age groups and are presented in Table 13.3. In all cases except the somatosensory-motor regions, this effect appeared to be slightly weaker in P12 pups than in P17 pups.

Limbic Thalamic/Cortical. The first LV ($p < 0.01$, 91.6% SSCB) indicated P12 design scores of [-.73, -.75, 1.5] and P17 design scores of [-1.4, -1.0, 2.4]. In both cases, the first two design scores, associated with PSA and PRF groups, are negative relative to the third (HC group). This indicates that PSA and PRF values are increased in P17 relative to the P12 pups in the orbital frontal cortex, centromedian nucleus, and medial dorsal nucleus. The design scores are slightly smaller in magnitude among P12 pups relative to P17, indicating that the effect is slightly greater in the P17 pups. The saliences associated with each are negative, indicating the increase in activity of P17 pups relative to P12.

Somatosensory/Motor. The design scores of [-1.0, -1.2, 2.2] for P12 pups and [-0.88, -0.85, 1.7] for P17 pups, associated with LV1 ($p < 0.01$, 95.4% SSCB), indicated the training effect. As indicated by the magnitudes of design scores, the effect is slightly smaller in the P17 pups relative to P12. The regions showing this effect were the anterior parietal cortex (-0.37, salience), ventral basal complex (-0.42), spinal trigeminal nucleus (-0.42), and vestibular nucleus (-0.45).

Brainstem. The gigantocellular nucleus and the interstitial nucleus of the medial longitudinal fasciculus were associated with P12 ([-.87, -2.2, 1.9]) and P17 ([-1.1, -0.9, 2.0]) design scores indicating training effects. This LV accounted for 96.5% of SSCB.

Cerebellum. LV1 ($p < 0.01$, 96.4% SSCB) indicated training effects in the cerebellar vermis, cerebellar hemisphere, and medial cerebellar nucleus. The effects were indicated by P12 design scores of [-.76, -.98, 1.7] and P17 design scores of [-1.1, -1.1, 2.2].

13.4.2.2 Comparison of PLS with Univariate Method

We compared the results from the PLS analysis of limbic cortical/thalamic regions to a repeated measures analysis followed by tests for simple effects of the same regions. A modified Bonferroni correction was used to control for multiple comparisons (Hochberg, 1988). A significant age by region effect ($F(9,324)=11.57$, $p < 0.05$) was found in the medial frontal, anterior cingulate, posterior cingulate, retrosplenial, perirhinal cortices, medial dorsal thalamus, and centromedian thalamus. These regions were increased in P17 pups relative to P12. FDG uptake in the orbital frontal cortex, medial dorsal thalamus and centromedian nucleus was increased in PSA and PRF pups relative to controls at both ages. This effect was significant at a 0.05 alpha level before Bonferroni correction, but not after correction. Thus, mean differences using the univariate repeated measures approach were limited to the regions showing an age effect.

13.4.3 Discussion

In summary, the cingulate and frontal cortices, amygdala, midline thalamic nuclei, cerebellum, and several brainstem regions demonstrated increases in activity between P12 and P17, independent of training. Second, analysis of residuals revealed a system of regions - the orbital frontal cortex, limbic thalamus, gigantocellular nucleus, the somatosensory system, and cerebellum - which demonstrated increases in uptake in the two trained groups relative to controls in both age groups. Age dependent training effects were found in the interpositus and medial cerebellar nuclei, and they showed increases in the PSA12 and PRF12 groups relative to controls. Aside from the anterior dorsal thalamus, which distinguished the PSA12 group from the PRF12 and HC12 groups, no other regions distinguished PSA and PRF groups at either age. Analysis of the Limbic/Thalamic-Cortical regions using a repeated measures ANOVA followed by tests for simple effects revealed similar age effects as found in the PLS analysis. However, the combined PLS analyses of raw and residualized data identified 3 regions demonstrating training differences (which were common to both age groups), while the ANOVA method revealed none. Corrections for multiple tests appeared to be quite conservative in the univariate case, accounting for the difference. Because PLS considers brain regions simultaneously, there is no need for such correction, resulting in more statistical sensitivity.

13.5 Seed PLS: Frontal Cortical Interactions and Development of Extinction

Despite the behavioral extinction rate difference between ages and P17 trained groups, there were no regions in the previous study that distinguished the older trained groups in the current analysis. One possibility is that covariance relationships involving particular regions could distinguish groups. We have shown previously that while absolute FDG uptake may not change in a region between groups,

the functional interactions between one region and another, as indicated by their correlated activity, may change with task or age (Nair and Gonzalez-Lima, 1999).

The frontal and limbic cortices are well known to be associated with adaptability of behavior to environmental changes. In particular the medial prefrontal (mPFC), orbitofrontal (OFC), and anterior cingulate (ACC) cortices appear to be quite important for switching response strategies, attaching reward value to events, or learning to avoid aversive situations (Woodward et al., 1999; Tremblay and Schultz, 2000; Poremba and Gabriel, 1997). Given the substantial structural and physiological maturation of these regions between P12 and P17, it was hypothesized that their functional maturation contributes to the age-related behavioral differences. Specifically, the regions inability to engage in concerted functional activity with other regions of the brain may be related to the behavioral impairments at the younger age.

The seed PLS analysis is uniquely suited for such an analysis as it can identify regions showing task related changes in covariance relationships with another ROI in a single omnibus step. For example, if 10 brain regions (out of a sampling of 60) were to change in their covariance relationships with a cortical region (the seed) across three different tasks, then the seed PLS would identify the effect and the particular pattern (e.g. high correlations in task 1, weaker correlations in task 2, uncorrelated in task 3). Such an analysis would be cumbersome using a univariate approach, as each correlation (between the seed ROI and the rest of the data set) would have to be tested across groups individually.

13.5.1 Methods

Seed PLS was used to identify regions that may be involved in functional relationships with the aforementioned cortical regions in an age- and training-dependent manner. The regions included in this analysis are presented in Table 13.4.

The seed PLS computation proceeds in the same manner as the design PLS, except that a set of ROIs replaces the original design contrasts. The seed PLS analysis identifies sets of regions whose covariances with another region (the 'seed ROI') change across tasks or are common to tasks (McIntosh and Gonzalez-Lima, 1998). The analysis begins by computing a cross correlation matrix between a vector of FDG uptake values for the seed ROI and another vector containing the values for the other regions of interest (ROIs) (Figure 13.4A). The same cross correlation is calculated for each experimental group; all correlation matrices are then stacked into a single matrix. SVD of the cross correlation matrix returns mutually orthogonal LVs - one describing the change in covariances across groups, and the other indicating regions most contributing to the effect - and singular values (Figure 13.4B). The magnitude of the weights of the seed LV indicate how well the regions covary with the seed LV saliences. In this case each SV indicates the proportion of the cross block correlation (% SSCB) accounted for by the covariation between the seed region and saliences.

Table 13.4 Regions of Interest

Medial Prefrontal Cortex	(mPFC)	Anterior Dorsal Thalamic Nucleus	(AD)
Orbitofrontal Cortex	(OFC)	Paratenial Nucleus	(PT)
Anterior Cingulate Cortex	(ACC)	Medial Habenula	(MHB)
Perirhinal Cortex	(PRH)	Lateral Habenula	(LHB)
Retrosplenial Cortex	(RSC)	Dorsal Raphe	(DR)
Posterior Cingulate Cortex	(PCG)	Pedunculopontine Tegmental Nucleus	(PTG)
Subiculum	(Sub)	Gigantocellular Nucleus	(GIG)
	CA1	Nucleus Accumben	(NAC)
	CA3	Caudate Putamen	(CPU)
Dentate Gyrus	(DG)	Globus Pallidus	(GP)
Basolateral Amygdala	(BLA)	Substantia Nigra	(SN)
Central Amygdala	(CEA)	Ventral Tegmental Area	(VTA)
Medial Amygdala	(MEA)	Anterior Parietal Cortex	(PAR)
Mediodorsal Thalamus	(MD)	Ventrobasal Complex	(VB)
Centromedian Nucleus, Thalamus	(CM)	Spinal Trigeminal Nucleus	(SPT)

13.5.2 Seed PLS Statistical Significance

The statistical significance of each LV pair was assessed via a permutation test of the singular value corresponding to the pair (McIntosh et al., 1996). The subject to group assignment for the seed ROIs was randomized, and the PLS recomputed. This was repeated 5000 times and the probability of a singular value greater than or equal to the original was computed. Those singular values with a probability of less than 0.01 were considered significant. The individual saliences were tested in the same way. Because all regions are considered simultaneously by the PLS analysis, there is no need to correct for multiple comparisons in the permutation tests.

13.5.3 Results

Permutation tests of singular values revealed that the first LV pair returned in each seed PLS analysis for the P17 groups was significant (p¡0.01). The percent SSCB for each of the three cortical regions were: mPFC, 80% ; OFC, 88%; ACC, 82%. In summary, the analyses identified 1) a pattern of covariance change involving the mPFC which distinguished the PSA17 group from the PRF17 and HC17 groups; 2) covariances involving the OFC which were relatively similar across groups but were slightly higher in PSA17 pups; and 3) an ACC covariance pattern which distinguished PSA17 and PRF17 pups from controls. The mPFC, OFC, and ACC seed LVs and those regions demonstrating significant saliences (according to permutation tests, $p < 0.01$) are presented in Figures 13.5, 13.6

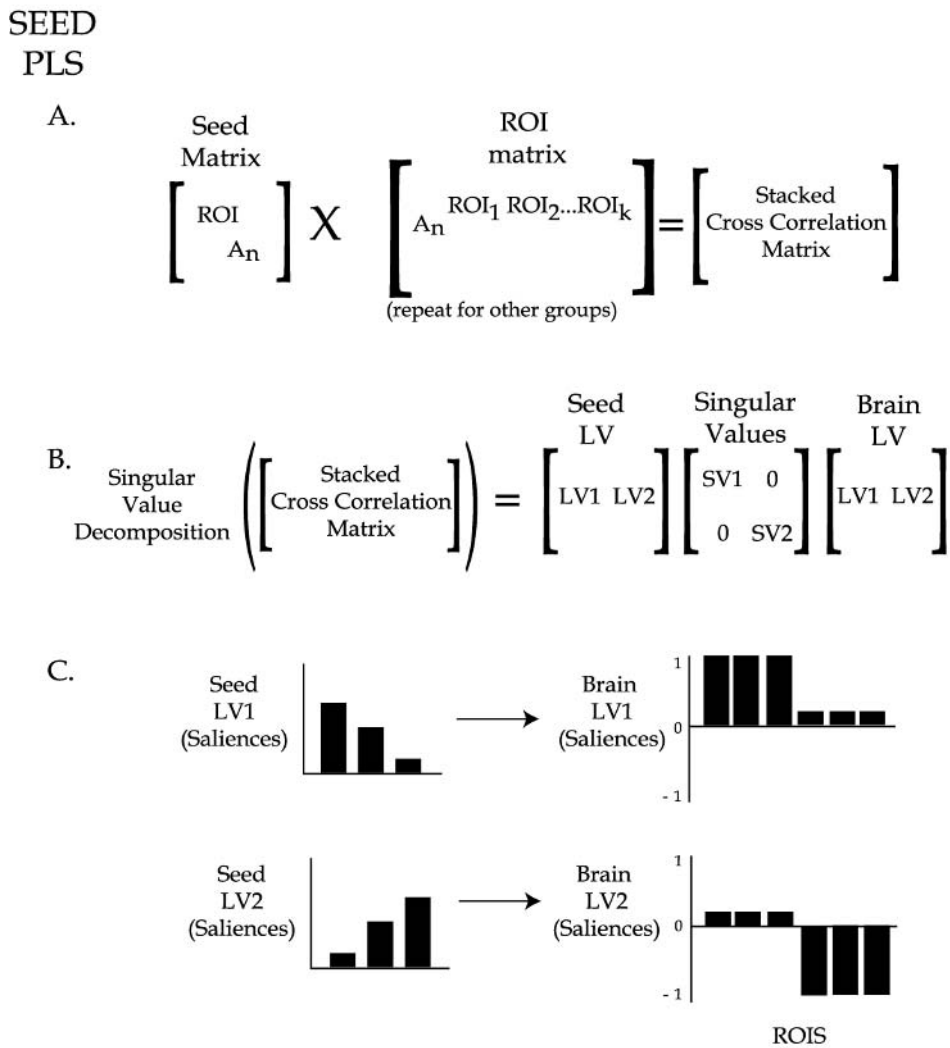


Figure 13.4 A) and B) are the same computations as in the design PLS, except that the design contrasts are replaced by a vector values for a single ROI. C) The effect indicates that the largest proportion of the cross-block covariation is accounted for by a change in covariances between the seed region and the first three ROIs, such that the greatest covariation occurs in group 1, less in group 2, and least in group 3. The next highest proportion is accounted for by a contrast such that group 3 demonstrates large, negative correlations across the last three regions; the magnitude of the same negative correlations are decreased in groups 1 and 2. Arrows indicate that the seed LVs show maximal covariance with their corresponding Brain LV.

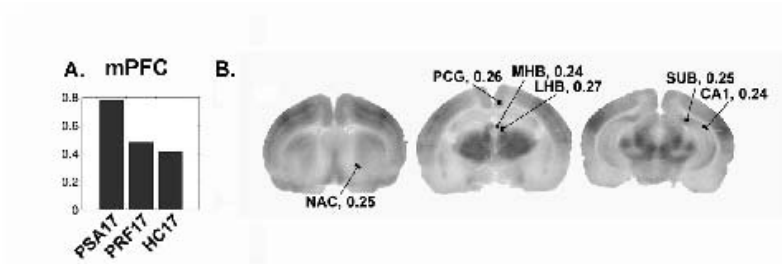


Figure 13.5 Seed PLS analysis of mPFC. (A) The graph of contrast vector for LV1 indicates that covariances between the mPFC were generally high across regions in PSA17, while they were less coupled in the PRF17 and HC17 groups. The regions significantly contributing to this effect according to the permutation tests are indicated in the autoradiographic images in (B). The value of the saliences are indicated next to each label. Note that all saliences are positive, indicating positive covariance relationships.

and 13.7, respectively. No significant LVs were returned by any of the seed PLSs performed for the P12 groups. Hence, while PLS identified functional interactions that distinguished or were common to the P17 groups, there were no significant covariance patterns associated with the P12 groups.

mPFC. The seed saliences associated with the mPFC were 0.77, 0.48, and 0.41 for PSA17, PRF17, and HC17 groups respectively. The regions significantly contributing to this pattern were the nucleus accumbens, posterior cingulate, medial habenula, lateral habenula, subiculum, and CA1 (the salience values associated with each region are presented in Figure 13.5B, next to each label). These results indicate that these six regions were most correlated with the mPFC in the PSA17 group relative to the PRF17 and HC17 groups.

OFC. The seed LV and salient brain regions associated with the OFC are presented in Figures 13.6A and 13.6B, respectively. The seed saliences were 0.69, 0.51, and 0.51 for PSA17, PRF17, and HC17 groups, respectively. The regions significantly contributing to this pattern were the anterior parietal cortex, anterior dorsal, medial dorsal, centromedian, and ventro-basal thalamic nuclei, posterior cingulate cortex, lateral habenula, ventral tegmental area, and gigantocellular nucleus. The higher salience in PSA17 pups indicates stronger covariances in this group. Saliences were similar between PRF17 and HC17 pups.

ACC. The seed saliences associated with the ACC were 0.74, 0.65, and 0.15 for PSA17, PRF17, and HC17 groups respectively (Figure 13.7A). The regions significantly contributing to this pattern were the medial amygdala, subiculum, CA1, CA3, perirhinal cortex, and the pedunculopontine nucleus (Figure 13.7B). Hence, while covariances between these regions and the anterior cingulate were similarly positive between PSA17 and PRF17 groups (although somewhat higher in the PSA17 pups), they were lower in the HC17 case.

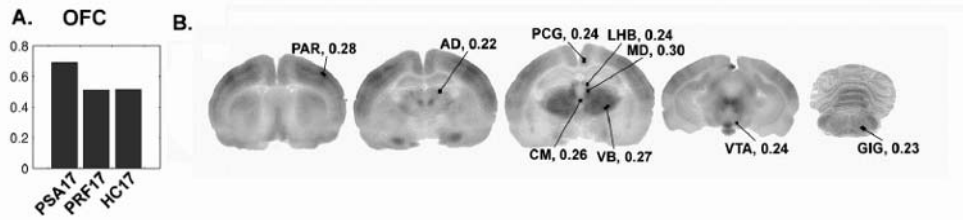


Figure 13.6 Seed PLS analysis of OFC covariances. (A) The graph of contrast vector for LV1 indicates that covariances between the OFC and sampled regions were generally high across all three P17 groups but slightly higher in the PSA17 pups. The regions significantly contributing to this effect according to the permutation tests are indicated in the autoradiographic images in (B). The value of the saliences are indicated next to each label.

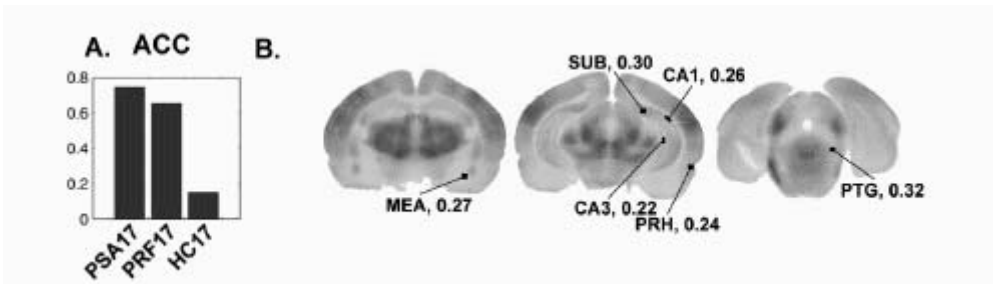


Figure 13.7 Seed PLS analysis of ACC covariances. (A) The graph of contrast vector for LV1 indicates that the regions were generally uncoupled in the HC17 pups relative to PRF17 and PSA17 pups. The regions significantly contributing to this effect according to the permutation tests are indicated in the autoradiographic images in (B). The value of the saliences are indicated next to each label.

13.5.4 Discussion

The results of the seed PLS analysis revealed covariance patterns describing training-related effects, as well as covariances common to all three groups in P17 pups. The first training-related change involved the mPFC, and indicated a system of regions that showed high correlations in the PSA17 group. Coupling between mPFC activity and the activities of the nucleus accumbens, subiculum, CA1, habenular nuclei, and posterior cingulate cortex were associated with the rapid PSA extinction behavior. The same set of regions showed positive correlations in the PRF17 groups, but were less coupled across the system of regions. They were largely uncoupled in the HC17 group. Hence, the recruitment of these regions changed depending on the particular experimental condition and appeared to be most related to the performance differences between P17 groups. PLS identified regions whose covariances with the OFC were similar between PRF17 and HC17 groups but slightly higher in the PSA17 groups. This effect is similar to the effect associated with the mPFC, although lesser in magnitude. In the case of the ACC, covariances

were higher in PSA17 relative to PRF17 groups but not to the extent found in the mPFC case. Covariances were generally low in the non-extinction HC groups. The regions associated with this particular common pattern of covariances during extinction behavior were the hippocampus, perirhinal cortex, medial amygdala, and pedunculopontine tegmental nucleus. The P17 seed PLS analyses implicate the mPFC, OFC, ACC, and their interactions with regions distributed throughout the rostro-caudal extent of the brain in supporting the differential extinction responding among the older pups. The rapid behavioral extinction of PSA17 pups may be related to covariances involving the mPFC and OFC. The similarity in ACC covariance relationships between PSA17 and PRF17 groups suggests they may be related to behavioral extinction processes common to both groups, not found in the HC pups. There were no dominant patterns of covariances that were similar or different between P12 groups, indicating that at the younger age, the three cortical regions are operating differently. In fact, their functional dissociation may underlie the behavioral differences found among the younger group.

A caveat, however is that the seed PLS considers the correlations between the seed ROI and the rest of the data set as a whole. The method identifies dominant patterns of changes, and so information regarding individual correlations is not emphasized in the computation. For example, it is possible that one region out of those sampled demonstrated a significant covariance change across P12 groups but was not identified by PLS since this was not a dominant pattern. If information regarding a particular region is desired, complementing PLS with a univariate analysis may be required.

13.6 Summary and Conclusions

The combined approach of the design and seed PLS contributed substantially to identifying age and task related activational and covariances changes in FDG uptake. While age-related changes in several regions were found, a common network of regions was activated due to extinction across trained P12 and P17 groups. The seed PLS analysis, however, suggests that superimposed on this network, is a frontal cortical network in P17 pups which may allow them to behave flexibly during extinction. This was one of the hypotheses regarding the brain functional activity at the two ages. Hence, the data-driven PLS approach lends support to this a priori hypothesis. The utility of the PLS approach resides in its identification of task or age related changes in a relatively efficient manner. The same information as found here would undoubtedly be difficult to obtain via traditional univariate approaches. PLS, in conjunction with univariate or other multivariate techniques such as principal components or independent components analyses would be a powerful approach for maximally extracting information regarding large scale brain functional systems related to learning and memory.

Acknowledgments

Supported by NIH grants RO1 NS37755 to F.G.L. and F31 MH11968 to H.P.N. We thank Dr. A.R. McIntosh for his statistical advice and assistance and Abram Amsel for providing his lab for the behavioral work. We also thank Jason Berndt and Douglas Barrett for their assistance on implementing the behavioral training, tissue processing, and imaging.

References

C. R. Almli and R. S. Fisher. Infant Rats: Sensorimotor ontogeny and effects of substantia nigra destruction. Brain Res. Bull. 2: 425-459, 1977.

F. L. Bookstein. Partial Least Squares: A dose-response model for measurement in the behavioral and brain sciences. Psycoloquy 5: 1-10, 1994.

H. T. Chugani, D. A. Hovda, J. R. Villablanca, M. E. Phelps, and W. F. Xu. Metabolic maturation of the brain: A study of local cerebral glucose utilization in the developing cat. J. Cereb. Blood Flow Metab. 11: 35-47, 1991.

J. M. Fuster. Distributed memory for short and long term. Neurobiol. Learn. Mem. 70: 268-270, 1998.

F. Gonzalez-Lima. Brain imaging of auditory learning functions in rats: studies with fluorodeoxyglucose autoradiography and cytochrome oxidase histochemistry. In: Advances in Metabolic Mapping Techniques for Brain Imaging of Behavioral and Learning Functions (F. Gonzalez-Lima, T. Finkenstadt, and H. Scheich (eds), 39-109, Kluwer Academic Press, Boston, MA, 1992.

Y. Hochberg. A sharper Bonferroni procedure for multiple tests of significance. Biometrics 75: 800-802, 1988.

B. Horwitz, T. T. Soncrant, and J. V. Haxby. Covariance analysis of functional interactions in the brain using metabolic and blood flow data. In: Advances in Metabolic Mapping Techniques for Brain Imaging of Behavioral and Learning Functions (F. Gonzalez-Lima, T. Finkenstadt, and H. Scheich (eds), 189-212, Kluwer Academic Press, Boston, MA, 1992.

C. Kennedy, G. D. Grave, J. W. Jehle, and L. Sokoloff. Changes in blood flow in the component structures of the dog brain during postnatal maturation. J. Neurochem. 19: 2423-2433, 1972.

W. Lilliquist, H. P. Nair, F. Gonzalez-Lima, and A. Amsel. Extinction after regular and irregular reward schedules in the infant rat: influence of age and training duration. Dev. Psychobiol. 34: 57-70, 1999.

A. R. McIntosh. Mapping cognition to the brain through neural interactions. Memory. 7: 523-548, 1999.

A. R. McIntosh, F. L. Bookstein, J. V. Haxby, and C. L. Grady. Spatial pattern analysis of functional brain images using partial least squares. Neuroimage. 3: 143-157, 1996.

A.R. McIntosh and F. Gonzalez-Lima. Large-scale functional connectivity in associative learning: interrelations of the rat auditory, visual, and limbic systems. J. Neurophysiol. 80: 3148-3162, 1998.

H. P. Nair, J. D. Berndt, D. Barret, and F. Gonzalez-Lima. Maturation of extinction behavior in infant rats: Large-scale regional interactions with medial prefrontal cortex, orbitofrontal cortex, and anterior cingulate cortex. J. Neurosci. 21, 4400-4407, 2001a

H.P. Nair, J.D. Berndt, D. Barrett, and F. Gonzalez-Lima. Metabolic mapping of brain regions associated with behavioral extinction in preweanling rats. Brain Res. 901, 141-153, 2001b.

H. P. Nair, T. Collisson, and F. Gonzalez-Lima. Postnatal development of cytochrome oxidase activity in fiber tracts in the rat brain. Dev. Brain Res. 118: 199-205, 1999.

H. P. Nair and F. Gonzalez-Lima. Extinction of behavior in infant rats: development of functional coupling between septal, hippocampal, and ventral tegmental regions. J. Neurosci. 19: 8646-8655, 1999.

A. Poremba and M. Gabriel. Amygdalar lesions block discriminative avoidance learning and cingulothalamic training-induced neuronal plasticity in rabbits. J. Neurosci. 17: 5237-5244, 1997.

R. A. Reyment, K. G. Joreskog, and L. F. Marcus. Applied Factor Analysis in the Natural Sciences. Cambridge University Press, Boston, MA, 1996.

L. Tremblay and W. Schultz. Reward-related neuronal activity during go-nogo task performance in primate orbitofrontal cortex. J. Neurophysiol. 83: 1864-1876, 2000.

D. J. Woodward, J. Y. Chang, P. Janak, A. Azarov, and K. Anstrom. Mesolimbic neuronal activity across behavioral states. Ann. N. Y. Acad. Sci. 877: 91-112, 1999.

Index