ABSTRACT

PREDICTIVE DECODING OF NEURAL DATA

by
Yaroslav O. Halchenko

In the last five decades the number of techniques available for non-invasive functional imaging has increased dramatically. Researchers today can choose from a variety of imaging modalities that include EEG, MEG, PET, SPECT, MRI, and fMRI.

This doctoral dissertation offers a methodology for the reliable analysis of neural data at different levels of investigation. By using statistical learning algorithms the proposed approach allows single-trial analysis of various neural data by decoding them into variables of interest. Unbiased testing of the decoder on new samples of the data provides a generalization assessment of decoding performance reliability. Through consecutive analysis of the constructed decoder's sensitivity it is possible to identify neural signal components relevant to the task of interest. The proposed methodology accounts for covariance and causality structures present in the signal. This feature makes it more powerful than conventional univariate methods which currently dominate the neuroscience field.

Chapter 2 describes the generic approach toward the analysis of neural data using statistical learning algorithms. Chapter 3 presents an analysis of results from four neural data modalities: extracellular recordings, EEG, MEG, and fMRI. These examples demonstrate the ability of the approach to reveal neural data components which cannot be uncovered with conventional methods.

A further extension of the methodology, Chapter 4 is used to analyze data from multiple neural data modalities: EEG and fMRI. The reliable mapping of data from one modality into the other provides a better understanding of the underlying neural processes. By allowing the spatial-temporal exploration of neural signals under loose

modeling assumptions, it removes potential bias in the analysis of neural data due to otherwise possible forward model misspecification.

The proposed methodology has been formalized into a free and open source Python framework for statistical learning based data analysis. This framework, PyMVPA, is described in Chapter 5.

PREDICTIVE DECODING OF NEURAL DATA

by
Yaroslav O. Halchenko

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Computer Science

Department of Computer Science

May 2009

APPROVAL PAGE

PREDICTIVE DECODING OF NEURAL DATA

Yaroslav O. Halchenko

Dr. Jason T. L. Wang, Dissertation Co-Advisor — Date
Professor of Computer Science, NJIT

Dr. Stephen J. Hanson, Dissertation Co-Advisor — Date
Professor of Psychology, Rutgers-Newark

Dr. Zoi-Heleni Michalopoulou, Committee Member — Date
Professor of Mathematical Sciences & Electrical and Computer Engineering, NJIT

Dr. David Nassimi, Committee Member — Date
Associate Professor of Computer Science, NJIT

Dr. Chengjun Liu, Committee Member — Date
Associate Professor of Computer Science, NJIT

Dr. Barak A. Pearlmutter, Committee Member — Date
Professor of Computer Science, Hamilton Institute, National University of Ireland, Maynooth

BIOGRAPHICAL SKETCH

Author: Yaroslav O. Halchenko

Degree: Doctor of Philosophy

Date: May 2009

Date of Birth: August 13, 1977

Place of Birth: Omsk, Russian Federation

Undergraduate and Graduate Education:

- Doctor of Philosophy in Computer Science,
  New Jersey Institute of Technology, Newark, NJ, USA, 2009
- Master of Science in Computer Science,
  University of New Mexico, Albuquerque, NM, USA, 2002
- Master of Science in Laser and Optoelectronic Engineering,
  Vinnytsya State University, Vinnytsya, Ukraine, 1999
- Bachelor's Degree in Laser and Optoelectronic Engineering,
  Vinnytsya State University, Vinnytsya, Ukraine, 1998

Major: Computer Science

Presentations and Publications:

J. Ramsey, S. J. Hanson, C. Hanson, Y. O. Halchenko, R. A. Poldrack, and C. Glymour. Six problems for causal inference from fMRI, In submission.

R. A. Poldrack, Y. O. Halchenko, and S. J. Hanson. Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychological Science*, In press.

M. Hanke, Y. O. Halchenko, P. B. Sederberg, and J. M. Hughes. *The PyMVPA Manual*. Available online at http://www.pymvpa.org/PyMVPA-Manual.pdf.

M. Hanke, Y. O. Halchenko, P. B. Sederberg, S. J. Hanson, J. V. Haxby, and S. Pollmann. PyMVPA: A Python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*, 7(1):37-53, Mar. 2009.

M. Hanke, Y. O. Halchenko, P. B. Sederberg, E. Olivetti, I. Fründ, J. W. Rieger, C. S. Herrmann, J. V. Haxby, S. J. Hanson, and S. Pollmann. PyMVPA: A unifying approach to the analysis of neuroscientific data. *Frontiers in Neuroinformatics*, 3 (3), 2009.

S. J. Hanson and Y. O. Halchenko. Brain reading using full brain support vector machines for object recognition: There is no face identification area. *Neural Computation*, 20 (2):486-503, February 2008.

S. J. Hanson, C. Hanson, Y. O. Halchenko, T. Matsuka, and A. Zaimi. Bottom-up and top-down brain functional connectivity underlying comprehension of everyday visual action. *Brain Struct Funct*, 212(3-4):231-44, 2007.

S. J. Hanson, D. Rebbechi, C. Hanson, and Y. O. Halchenko. Dense mode clustering in brain maps. *Magn Reson Imaging*, 25(9):1249-62, 2007.

Y. O. Halchenko, S. J. Hanson, and B. A. Pearlmutter. Multimodal integration: fMRI, MRI, EEG, and MEG. In L. Landini, V. Positano, and M. F. Santarelli, editors, *Advanced Image Processing in Magnetic Resonance Imaging*, Signal Processing and Communications, chapter 8, pages 223-265. Dekker, 2005. ISBN 0824725425.

S. J. Hanson, T. Matsuka, C. Hanson, D. Rebbechi, Y. O. Halchenko, A. Zaimi, and B. A. Pearlmutter. Structural equation modeling of neuroimaging data: Exhaustive search and Markov chain Monte Carlo. In *Human Brain Mapping*, 2004.

Y. O. Halchenko, S. J. Hanson, and B. A. Pearlmutter. Fusion of functional brain imaging modalities using l-norms signal reconstruction. In *Proceedings of the Annual Meeting of the Cognitive Neuroscience Society*, San Francisco, CA, 2004.

Y. O. Halchenko, B. A. Pearlmutter, S. J. Hanson, and A. Zaimi. Fusion of functional brain imaging modalities via linear programming. *Biomedizinische Technik (Biomedical Engineering)*, 48(2):102-104, 2004.

L. I. Timchenko, Y. F. Kutaev, A. A. Gertsiy, Y. O. Halchenko, L. V. Zahoruiko, and T. Mansur. Method for image coordinate definition on extended laser paths. In S. B. Gurevich, Z. T. Nazarchuk, and L. I. Muravsky, editors, *Optoelectronic and Hybrid Optical/Digital Systems for Image and Signal Processing*, volume 4148:1, pages 19-26. SPIE, 2000.

L. I. Timchenko, Y. F. Kutaev, A. A. Gertsiy, L. V. Zahoruiko, Y. O. Halchenko, and T. Mansur. Approach to parallel-hierarchical network learning for real-time image sequence recognition. In J. W. V. Miller, S. S. Solomon, and B. G. Batchelor, editors, *Machine Vision Systems for Inspection and Metrology VIII*, volume 3836:1, pages 71-81. SPIE, 1999.

L. I. Timchenko, A. A. Gertsiy, L. V. Zahoruiko, Y. F. Kutaev, and Y. O. Halchenko. Pre-processing of extended laser path images. In *Industrial Lasers and Inspection: Diagnostic Imaging Technologies and Industrial Applications*, Munich, Germany, June 1999. EOS/SPIE International Symposium. [3827-26].

L. I. Tymchenko, J. Scorukova, S. Markov, and Y. O. Halchenko. Image segmentation on the basis of spatial connected features. In *Visnyk VSTU*, volume 4, pages 39-43. VSTU University Press, Vinnytsya, Ukraine, 1998. In Ukrainian.

T. B. Martinyuk, A. V. Kogemiako, and Y. O. Halchenko. The model of associative processor for numerical data sorting. In *Visnyk VSTU*, volume 2, pages 19-23. VSTU University Press, Vinnytsya, Ukraine, 1997. In Ukrainian.

L. I. Tymchenko, J. Scorukova, J. Kutaev, S. Markov, T. Martynuk, and Y. O. Halchenko. Method spatial connected segmentation of images. In *The Third All-Ukrainian International Conference Ukrobraz*, Kijiv, Ukraine, November 1996. In Ukrainian.

Моим любимым старичкам:
Петру и Клавдии Курасановым

LIST OF TERMS

Abbreviation	
ANOVA	Analysis of Variance
BCI	Brain Computer Interface
BEM	Boundary Elements Method
BMA	Bayesian Model Averaging
BOLD	Blood Oxygenation Level Dependent
DC	Direct Current
DECD	Distributed ECD
ECD	Equivalent Current Dipole
EEG	Electroencephalography
EIT	Electrical Impedance Tomography
EMF	Electromotive Force
E/MEG	EEG and/or MEG
EM	Expectation Maximization
EMSI	Electromagnetic Source Imaging
EPI	Echo Planar Imaging
ER	Event-related
ERF	Event-related Magnetic Field (in MEG)
ERP	Event-related Potential (in EEG)
FDA	Food and Drug Administration
FDM	Finite Difference Method
FEM	Finite Elements Method
FFA	Fusiform Face Area
FG	Fusiform Gyrus
FLIRT	FMRIB's Linear Image Registration Tool
FMRI	Functional MRI
FMRIB	Oxford Centre for Functional MRI of the Brain
FOSS	Free and Open Source Software
FSL	FMRIB Software Library
GLM	General Linear Model
GPR	Gaussian Process Regression
HbO	Oxy-hemoglobin
HbR	Deoxy-hemoglobin
HbT	Hemoglobin (Total Count)
HR	Hemodynamic Response
HRF	HR Function
ICA	Independent Component Analysis
IED	Interictal Epileptic Discharge
LCMV	Linearly Constrained Minimum Variance
LFP	Local Field Potential

LO	Lateral Occipital
LP	Linear Programming
LTIS	Linear Time Invariant System
MAP	Maximum a Posteriori
MANOVA	Multivariate Analysis of Variance
MCMC	Monte-Carlo Markov Chain
MEG	Magnetoencephalography
MFG	Middle Frontal Gyrus
ML	Machine Learning
MR	Magnetic Resonance
MRI	MR Imaging
NIRS	Near-Infrared Spectroscopy
NMR	Nuclear Magnetic Resonance
PCA	Principal Component Analysis
PCC	Posterior Cingulate Cortex
PD	Proton Density Imaging
PDF	Probability Density Function
PET	Positron Emission Tomography
PCUN	Precuneous
PHG	Parahippocampal Gyrus
PLS	Partial Least-Squares
PyMVPA	Python Multivariate Pattern Analysis Toolbox
pSTS	Posterior Superior Temporal Sulcus
RFE	Recursive Feature Elimination
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SMLR	Sparse Multinomial Logistic Regression
SNR	Signal-to-noise Ratio
SOBI	Second Order Blind Identification
SPECT	Single Photon Emission Computed Tomography
SPM	Statistical Parametric Mapping
SQUID	Superconducting Quantum Interference Device
STS	Superior Temporal Sulcus
STG	Superior Temporal Gyrus
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SVR	Support Vector Machine Regression
TOFC	Temporal Occipital Fusiform Cortex
TR	Time of Repetition (in fMRI)
VEP	Visual Evoked Potential
WMN	Weighted Minimum Norm

Symbol

$\mathbf{0}$	Zero matrix of appropriate dimensionality
$\mathbf{C}$	Covariance matrix
$\mathbf{F}$	BOLD fMRI data matrix ($N \times U$)
$\mathbf{F}^i$	Spatial filter matrix for the i-th dipole ($M \times L$)
$\mathbf{G}$	General E/MEG lead matrix
$\mathbf{I}_n$	Identity matrix ($n \times n$)
K	Number of simultaneously active voxels
$\mathbf{K}$	Matrix of correlation coefficients
L	Number of orthogonal axes for dipole moment components, $L \in \{1, 2, 3\}$
M	Number of EEG/MEG sensors, *i.e.*, spatial resolution of low spatial resolution modality
N	Number of voxels, *i.e.*, spatial resolution of high spatial resolution modality (fMRI)
$\mathbf{Q}$	General dipole sources matrix ($LN \times T$)
$\tilde{\mathbf{Q}}$	Dipole sources strength matrix ($N \times T$)
T	Number of time points of high temporal resolution modality (EEG, MEG)
U	Number of time points of low temporal resolution modality (fMRI)
$\mathbf{X}$	General E/MEG data matrix; can contain EEG or/and MEG data ($M \times T$)
$\mathcal{G}$	General E/MEG lead function, incorporating information for EEG or/and MEG
Θ	Dipole sources orientation matrix ($LN \times T$)
σ	Deviation
$\mathcal{T}$	General time-frequency transformation function
ν	Variance

Function

$\mathbf{M}^\top$	Matrix transpose
$\mathbf{M}^+$	Generalized matrix inverse (pseudo-inverse)
$\operatorname{null} \mathbf{M}$	The *null space* of $\mathbf{M}$, the set of vectors $\{\mathbf{x} \mid \mathbf{M}\mathbf{x} = 0\}$
$\operatorname{diag} \mathbf{M}$	The diagonal matrix with the same diagonal elements as $\mathbf{M}$

GLOSSARY OF TERMS

Chunk
A group of data samples which are independent from the other samples. This information is important in the context of a cross-validation procedure. Different chunk ids could correspond to different subjects, experimental runs, *etc.*

Cross-validation
A technique for assessing how the results of a statistical analysis will generalize to an independent dataset. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice (Description from the Wikipedia).

Dataset
A combination of data samples and related attributes (*e.g.*, samples labels and chunks, channel ids, *etc.*).

Label
A value associated with a data sample (or multiple samples). It can correspond to a certain category, experimental condition or, in case of a regression problem, to a variable of ordered numbers. Labels therefore define the model that a learner has to learn, and are used when cross-validating the performance of the learner.

Learner
A method capable of devising a mapping from the space of data samples into the space of labels. If labels belong to a non-ordered set, learner is called a *classifier*. If labels are ordered and there is a corresponding distance metric, *regression*.

Machine Learning
A field of Computer Science which aims at constructing methods, such as learners, to integrate available knowledge extracted from the existing data.

Mass-univariate
An analysis scheme when a univariate method applied to all observables (variables) one observable at a time.

Multivariate
Having or involving a number of independent mathematical or statistical variables (Definition from the Merriam-Webster Online Dictionary). In statistics it describes a collection of procedures which involve observation and analysis of more than one statistical variable at a time.

Neural Data Modality
A reflection of neural activity collected using some available instrumental method (*e.g.*, EEG, MEG, fMRI, *etc.*).

Sensitivity
: A score assigned to a particular feature with respect to its impact on the learner's decision.

Statistical Learning
: A field of Computer Science which aims at exploiting statistical properties of the data to construct reliable learners (hence it is related to Machine Learning), and to assess their convergence and generalization performances.

Univariate
: Characterized by or depending on only one random variable (Definition from the Merriam-Webster Online Dictionary). In mathematics univariate refers to an expression, equation, function or polynomial of only one variable. Objects of any of these types but involving more than one variable may be called multivariate. In statistics, this term is used to distinguish a distribution of one variable from a distribution of several variables.

ACKNOWLEDGMENT

A dissertation is one of the most important milestones in the life of any Ph.D. student. As such, it forms and evolves together with a student, from raw marble into what one hopes is a fine piece of scientific art. As any artwork it cannot stand alone, uninfluenced by the outside world, thus it looks for the inspiration and advice from it's surrounding, which, in turn, plays an important role in the shaping and molding, aspiring for perfection of the said work.

My lengthy journey toward a Ph.D. degree began at the University of New Mexico under the supervision and guidance of Dr. Barak Pearlmutter. He became the ultimate source of inspiration for me, as he constantly encouraged me to grasp knowledge and to dive into the fascinating field of computer science and brain imaging. Aside from purely academic interests, he was the one who introduced me to the Debian GNU/Linux operating system, which I became the proud developer of, during my tenure as a graduate student.

Fine-grained polishing of my research work was delicately done by Dr. Stephen Hanson. I would like to thank him for the continuous stream of interesting projects to tackle and the supply of coffee beans to power up the gray matter early in the mornings. For his patience, advice and stimulating debates that helped me navigate my way to completion of this thesis work. For furnishing my trips to scientific conferences around the globe and setting up the meetings and collaborations with scientists working on the fascinating research.

Also, my scientific journey could not be accomplished without the often rescuing, and in other cases simply encouraging help from Dr. Jason Wang who was always there to help me find the answers to my questions. I would also like to thank the members of my committee Dr. Chengjun Liu, Dr. Eliza Michalopoulou, and Dr. David Nassimi for sharing with me the load of completing my dissertation, and for their skillful advice.

Although the universities were helping me to reach my final destination, I know, that it would be impossible to accomplish the mission without constant support and encouragement from my family: My parents Larisa Petrovna Halchenko and Ruslan Grigoryevich Nikolenko, my dear, loving and always supportive wife Yuliya. My daughters Alisa and Zabava who rushed to this world to join their mom in helping me grind the marble of science by providing additional joy and introducing me to another dimension of happiness. A special thanks to our favorite family pet – a cat who thought she was a mini-lion – Anfyska for annoying me a lot, but always leaving room for more. Unfortunately she passed on last year and is sorely missed.

Any journey is easier when you travel together, hence I want to express my gratitude to Mr. Michael Hanke, once a geographically distant but similar minded member of the Debian GNU/Linux community, now a collaborator and a close friend. I would like to thank him for his endless encouragement, ideas, data, and code sharing, and very fruitful joint research venue. Also I want to express my sincere regards to the other members of our PyMVPA team – Per Sederberg and Emanuele Olivetti – for the joy of sharing ideas, and sparing rare lonely evenings in the interesting and fertile discussions and article compositions. I also thank every member of RUMBA research laboratory at Rutgers-Newark for plentiful and joyful discussions and their help on a variety of occasions. A special "thank you" goes out to Dr. Catherine Hanson for her persistent helpful support.

It would be hard to arrive at the destination and furnish any interesting and meaningful results without having any interesting data for the analysis. Hence I want to thank everyone who shared their collected neural datasets: Dr. James V. Haxby, Dr. Christoph S. Herrmann, Dr. Artur Luczak, Dr. Kenneth D. Harris, and Dr. Ingo Fründ.

To conclude, I would like to thank all my friends and colleagues thanks to whom this journey reached its destination and National Science and McDonnel Foundations for providing financial support to empower my progress.

TABLE OF CONTENTS

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

> *The average Ph.D. thesis is nothing but the transference of bones from one graveyard to another.*
>
> – Frank J. Dobie *"A Texan in England"*, 1945

The brain is a complex, massively-parallel computing system composed of billions of neurons that it uses to extract, process, and store information about the physical world. The brain is capable of carrying out complex behavioral and cognitive tasks that range from walking to writing a Ph.D. dissertation. It does this through complex neural networks that recode the physical data available in the world into neural data, that is, it performs the information processing necessary for biological organisms to function in the physical world. The mechanisms that underlie information processing, the specialization of network function, and the interactions among different networks are issues of considerable interest and debate in the neuroscience community.

Neuroscientists can use existing technology to record the epiphenomena associated with neural activity and to extrapolate from this data the mechanisms underlying cognitive processing. These neural data represent large volumes of discrete signals which are complex, noisy, and non-stationary; a rather gross reflection of neural activity across spatial and temporal extents. For this reason, the use of advanced signal-processing methods, although underutilized in neuroscience, could potentially resolve many debates about brain function.

One debate in neuroscience involves the identification of functional specificity in the brain and the mapping of *specific* modules to a set of cognitive or perceptual processes. Another debate concerns the nature of the interaction among brain modules that facilitate *data* flow and data processing while engaged in cognitive tasks.

Current signal-processing methods used to analyze neural data are geared primarily toward *identification*. These methods are used to map functional modules or spatio-temporal components of the brain signal *specific* to some form of information processing, for example early visual processing. Such methods are often able to detect brain areas engaged in a given mental task, but make limited use of neural data, and consequently are often inefficient. Moreover, it remains an open question whether such methods can be used to claim *specificity* of a given area for a unique aspect of information processing (*e.g.*, recognition of the human face).

Because most of the signal-processing methods that are used in neuroscience studies are univariate, any particular basic unit in the temporal and/or spatial dimension of the signal is considered to be independent of any other. Such basic units correspond to a single channel of acquisition in space (*e.g.*, in fMRI), or to a specific temporal offset from stimulus presentation in the temporal domain (*e.g.*, in EEG). Since each basic unit carries a very noisy and information-laden signal, univariate methods are capable of detecting only large gross level effects of neural activation. This feature of univariate approaches requires certain pre-processing steps that necessarily ignore some information embedded in the signal.

Several limitations exist when using univariate methods to analyze neural data. For example, it is often necessary to average registered neural responses over a large number of trials to boost the signal-to-noise ratio. This averaging obfuscates the variance in signals that are not precisely locked to the onset of the stimulus trials. Another problem is that the independence assumption of univariate approaches essentially ignores potential covariance between neighboring or distant units. A third problem is that *identification* approaches do not account for various sources of signal variance, nor do they address the *causal* structure of information processing in the brain. This feature of univariate methods makes them inherently unable to provide adequate mapping between a stimulus context and a given brain state. These problems make univariate methods very limited in their

ability to either provide insight into the specific nature of the encoding in each module, or to assess the *similarity* among stimulus conditions. Finally, it should be noted that conventional methods do not promote model testing, which makes it difficult to assess the reliability of the findings.

Multivariate methods do exist that can hypothetically account for all the data. Unfortunately most of the classical multivariate methods (*e.g.*, multivariate analysis of variance or MANOVA) require a large number of data samples to reliably assess the underlying structure of the signal. Because neural data has high dimensionality, the number of samples that are available is limited. This fact makes it hard (if not impossible) to use classical multivariate methods.

In the past decade though, in the domain of statistical learning theory, new methods have been developed. These methods often do not require large number of samples to achieve reasonable performance. Moreover, they inherently provide adequate decoding capabilities together with promoting model testing. These features of statistical learning methods make it possible to develop new strategies for the analysis of neural data. The goal of this dissertation is the construction and analysis of methods that can be used to decode brain signals and that will further our understanding of the mapping between brain state and the experimental conditions.

Learning theory methods will be of particular interest in that they are able to decode brain states, and to discover functional structures of the brain signal at a level of detail univariate approaches are not able to provide. These methods are used to examine *identification*, *specificity*, *causality*, and *similarity* features of neural data from the level of neurons to the level of interacting neural systems. Thus, learning theory methods are well suited to the task of revealing the distributed functional networks underlying perceptual and cognitive processing.

This dissertation will also provide a new methodology for conjoint multi-modal data analysis. The proposed method provides reliable mapping of the brain state reflected

in one acquired data modality (EEG) into that of another data modality (fMRI). Because data in both modalities are convoluted with modality-specific inclusions (*e.g.*, instrumental noise, modality-specific neuro-physiological processes), the reliable mapping between modalities should allow the recovery of any common structure between them, *i.e.*, the effective neuronal activity.

The following sections of this chapter will overview existing neural data modalities and provide further reasoning for the exploration of decoding methods using unimodal and multimodal analysis of brain data.

1.1 Neuroscience Overview

Neuroscience is a field of research devoted to the study of the nervous system. The brain (see Figure 1.1) receives the greatest attention from neuroscientists inasmuch as it constitutes the main functional unit of the nervous system. It controls all parts of the body, dictates patterns of behavior, empowers thoughts, and provides the mechanism to acquire, encode, and store knowledge about the world. Hence, brain function research is relevant for anyone with intellectual interest in human behavior, whether that interest is driven philosophically, experimentally, or clinically. Increasing knowledge about brain function also has a pragmatic aspect. Improving the techniques used to decode brain signals can make it easier to ”read minds” or to increase the independence of handicapped individuals through brain-computer interfaces.

The brain of a human is composed of a network of neurons and other supportive cells (*e.g.*, glial cells) that number in the tens of billions. Each neuron on average has thousands of connections to other neurons which allows them to form functional networks, each responsible for specific tasks from simple reception and transmission of information to much more complex processes underlying human behavior. Thus, neuroscience can be

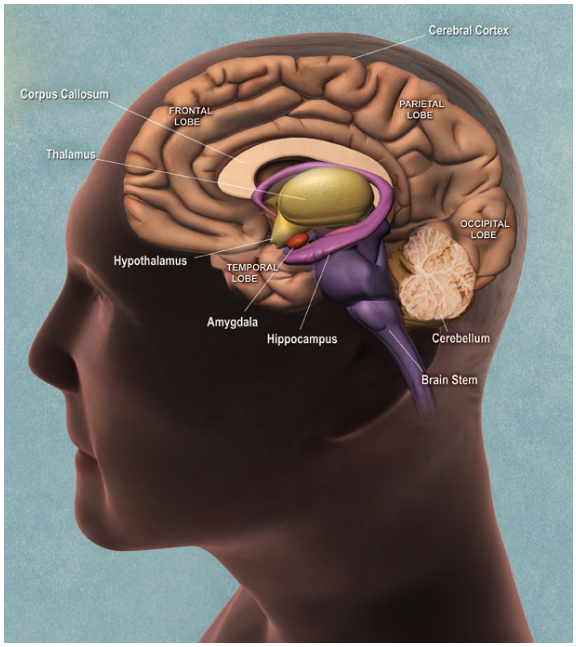


Figure 1.1 Side view of the human brain. Only right hemisphere is visualized.

studied at many different levels (as depicted in Figure 1.2), ranging from the molecular level to that of complex neural systems.

The most basic level of investigation in neuroscience happens at the cellular level, where the fundamental questions focus on the mechanisms underlying the ability of neurons to process signals physiologically and electrochemically. Researchers investigate how signals are processed by the dendrites, somas, and axons, and how neurotransmitters and electrical signals are used to conduct information between neurons. At a higher level, neuroscientists focus on networking of neurons including how these networks are formed. At this level of investigation, the goal is to model how particular networks are used to produce the physiological functions that range from simple reflexes to the complex processing involved in learning and memory. Research that is specifically aimed at understanding the role of complex systems and networks in cognitive behavior is designated as “cognitive neuroscience”.

While behavioral experiments allow scientists to uncover high-level aspects of human performance and behavior, they are incapable of uncovering the complexities

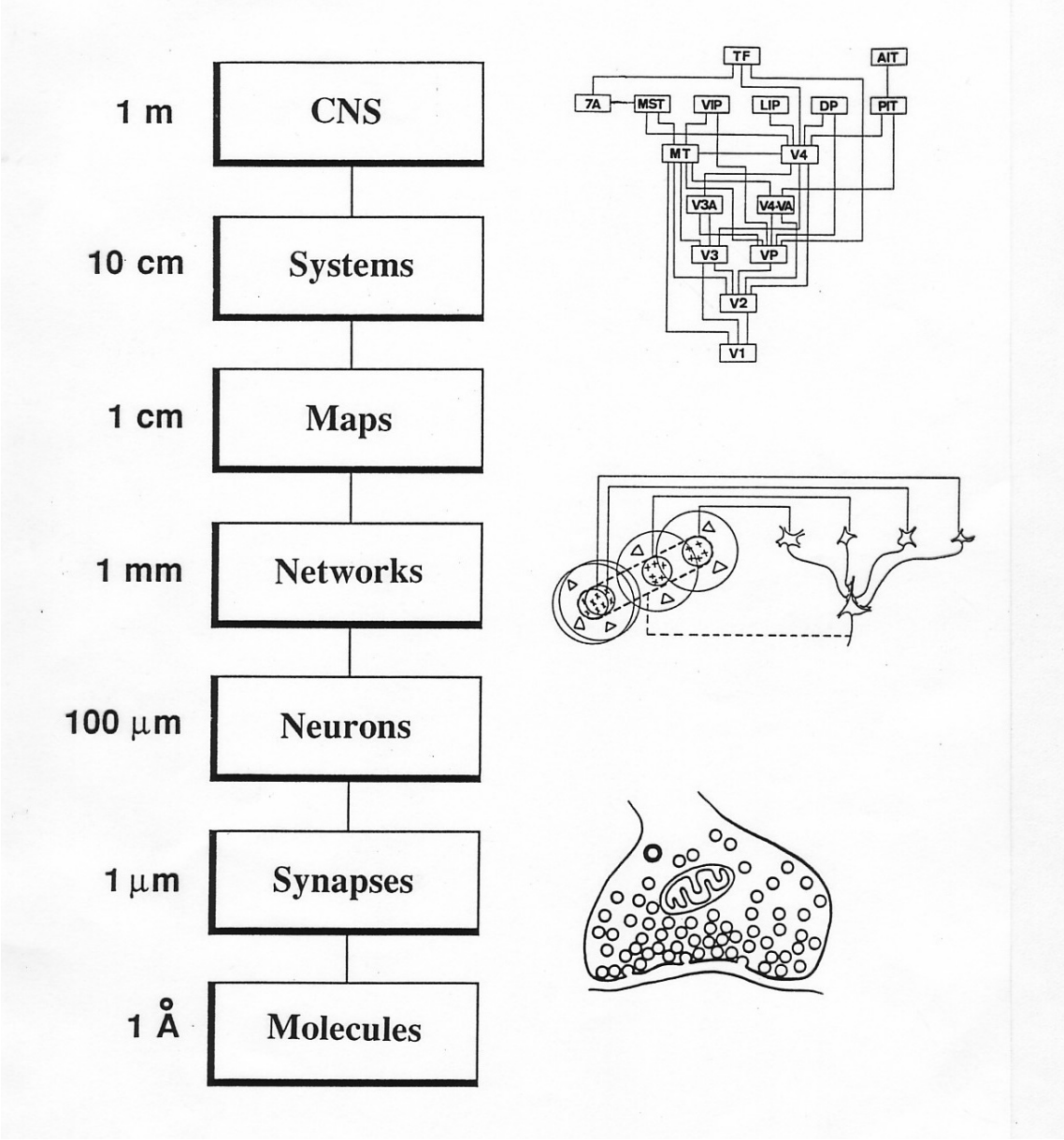


Figure 1.2 Different levels of human brain investigation [1].

involved in brain activation. Analysis of brain function must rely on data derived from the neural activity in the living brain.

One possible approach to understanding brain function is the use of simulation. There is an ongoing project "Blue Brain" [2], which aims to simulate the human brain down to the molecular level using existing neural models and available high-performance computing facilities. Thus far they have succeeded only in simulating a single rat neocortical column. This structure can be considered the smallest functional unit of the neocortex[1]. Such a column is about 2 mm tall, has a diameter of 0.5 mm, and contains only about 60,000 neurons in humans. Rat neocortical columns are very similar in structure but contain only 10,000 neurons (and 108 synapses). Despite this advancement, modeling of the whole human brain remains an ambitious and distant goal and even if possible would not provide direct answers about the higher-level functional structure of the brain.

The alternative approach available to neuroscientists interested in the exploration of brain function is the analysis of the neural signals of the brain. Unfortunately, at this point in time there is no single technology advanced enough to allow a "snapshot" of the complete neural system at either a molecular or neuronal level. Even if there were such technology available a format for precisely describing the billions of neurons with trillions of connections among them does not exist. However, even overcoming these technological problems would not magically resolve the debates in neuroscience. What is required are new data analysis approaches, which are able to extract relevant information from huge amounts of data. Moreover, these analysis methods would need to be customized for the level of investigation being followed (*i.e.*, cellular, network) as the questions being asked are quite different. It is for this reason that researchers interested in a given level of

[1]Neocortex is the part of the brain thought to be responsible for higher functions such as conscious thought

investigation have developed specialized analysis methods to capture and process neural activity at temporal and spatial scales of interest.

Research of human brain function presented a particular challenge in that a non-invasive[2] means of assessing the characteristics of neuro-physiological processes inside the brain was necessary. Invasive techniques, such as extracellular and intracellular recordings, which have been useful in describing neural functioning in animals, are not appropriate for the study of humans. Invasive techniques in humans are limited to those ancillary to medically required surgery such as that used to treat epilepsy. By adopting certain assumptions about the nature of neuronal signals, neuronal activity can be measured and modeled as biomedical signals. These signals can be obtained through several different types of non-invasive brain imaging modalities such as electroencephalography (EEG), magnetoencephalography (MEG), nuclear magnetic resonance (NMR) imaging (MRI)[3], positron emission tomography (PET), and near-infrared spectroscopy (NIRS).

These non-invasive neural data modalities can be categorized as either passive or active. The purpose of the passive modalities (EEG and MEG) is to register ambient environment changes which are caused by neuronal processes inside the brain. Active modalities (such as MRI, PET and NIRS) create a controllable environment which changes under influence of underlying neuronal and other physiological processes. Therefore, active modalities usually do not register neuronal activity directly, but rather capture changes caused by the neuronal activity, *e.g.*, consumption of contrast agents, blood oxygenation, or changes of blood flow. Captured brain signals by either passive or active modalities are usually non-stationary signals distorted by noise and signal interference. Moreover, they possess characteristics specific to the technique (modality)

---

[2]From WordNet (r) 2.0 (August 2003) [wn]: noninvasive adj: relating to a technique that does not involve puncturing the skin or entering a body cavity [ant: invasive]

[3]The term MRI generally substituted NMR so that the public could more easily adopt a term for an imaging modality without the word ”nuclear” in it

which was used to acquire it, so it is crucial to have a clear understanding of modality's nature when performing advanced signal analysis.

The oldest modality is EEG. It has been widely used in research and clinical studies since the mid-twentieth century. Although Richard Caton (1842–1926) is believed to have been the first to record the spontaneous electrical activity of the brain, the term EEG first appeared in 1929 when Hans Berger, a psychiatrist working in Jena, Germany, announced to the world that "it was possible to record the feeble electric currents generated on the brain, without opening the skull, and to depict them graphically onto a strip of paper." The first SQUID[4]-based MEG experiment with a human subject was conducted at MIT by Cohen [3] after his successful application of Zimmerman's SQUID sensors to acquire a magneto-cardiogram in 1969. EEG and MEG are closely related due to electro-magnetic coupling, and term E/MEG will be used to refer generically to either EEG, MEG, or both altogether. Although EEG and MEG are related, there are some subtle differences among them [4]. Both E/MEG provide high temporal resolution (milliseconds) but have a major limitation: the location of neuronal activity can be hard to pin-point with confidence. Localization of neural activity from E/MEG data is usually referred to as *electromagnetic source imaging* (EMSI) and has been a challenging area of research for the last couple of decades. That is because such techniques measure the signal outside of the head, that are created by the super-imposition of electromagnetic fields produced by activity at the level of neurons and neural networks. Therefore, in order to obtain even a coarse spatial localization of the activated neural populations an ill-posed inverse problem has to be solved.

Unlike E/MEG, MRI modality is inherently capable of providing an *in vivo* view of brain structure and function at the level of basic neural networks. The Nuclear Magnetic Resonance (NMR) effect was independently and simultaneously discovered by Felix Bloch and Edward Purcell in 1946. As a consequence of this important discovery,

[4]Superconducting Quantum Interference Device

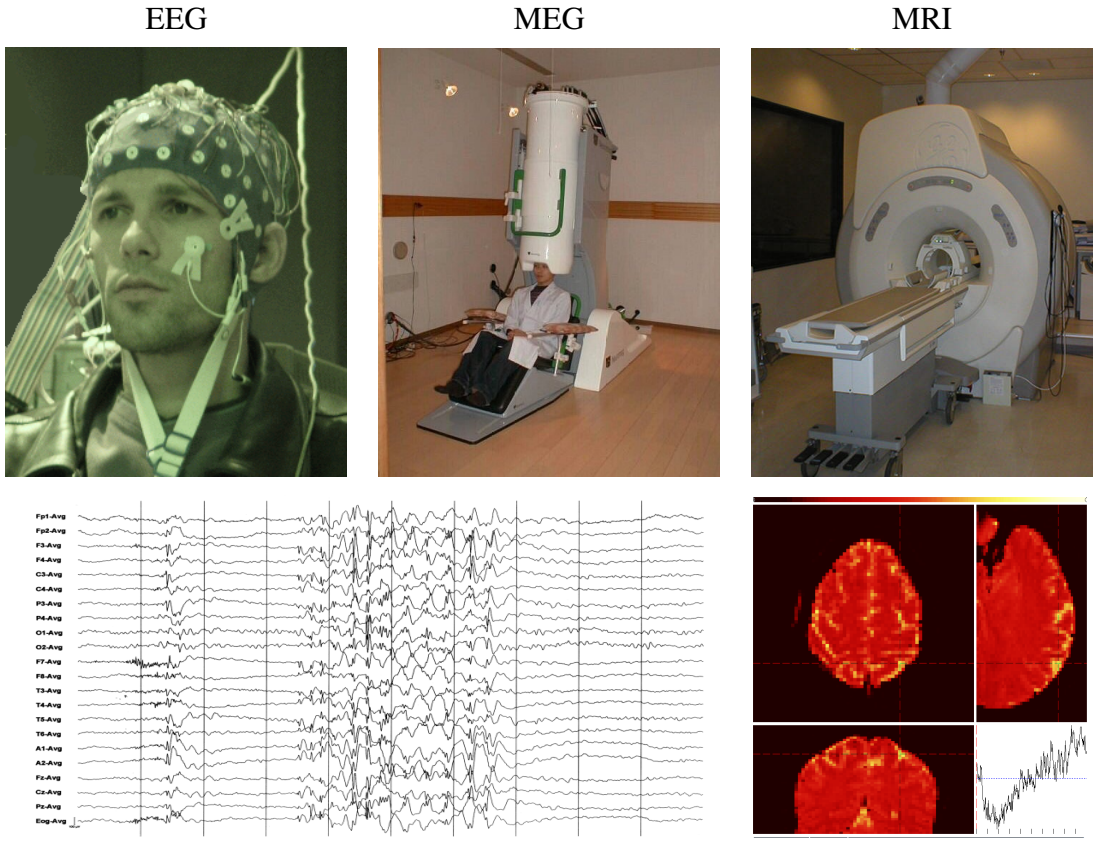


Figure 1.3 Non-invasive functional brain imaging equipment: from simple EEG to expensive MR. Top row shows the typical view of the used equipment, and bottom row visualizes typical data.

they shared a Nobel Prize in Physics in 1952. In 1970, Raymond Damidian came to the conclusion that the key to MR imaging (MRI) is the structure and abundance of water in the human body. It was Paul Lauterbur in 1973, however, who implemented the concept of tri-plane gradients used for exciting selective areas of the body. P. Lauterbur along with Peter Mansfield were awarded a Nobel Prize in Physiology and Medicine in 2003 for the invention of MRI, which made a huge impact on medical imaging.

Since their first appearance, MRI techniques have been evolving. Today image intensity observed in MR images can be determined by various tissue contrast mechanisms such as proton density, T1 and T2 relaxation rates, diffusive processes of proton spin dephasing, loss of proton phase coherence due to variations of magnetic

susceptibility in the tissue. Although MRI is capable of detecting transient or subtle changes in the magnetic field in the cortical tissue caused by neuronal activation [5, 6], the direct application of MRI for capturing functional activity remains limited due to a very low signal-to-noise ratio (SNR). For this reason, MRI is often labeled *anatomical*.

It was toward the end of the 19th century, when Charles Roy and Charles Sherrington [7] provided the first evidence supporting the connection between neuronal activity and cerebral blood flow. One hundred years after the MRI technique had received attention in anatomical studies, Ogawa et al. [8] it was discovered that MRI can reflect blood deoxygenation using T2*contrast. This finding laid down a framework for functional brain imaging using MRI [9, 10, 11] based on capturing blood oxygenation level-dependent (BOLD) signal without the use of any reactive agents; and made functional MRI (fMRI) the first truly non-invasive functional brain imaging modality able to provide rich spatial information. The typical spatial resolution of BOLD fMRI in human research is 1–3 mm, which is sufficient to provide information about neural activity at the level of neural networks. One drawback is that the hemodynamic (blood related) response lags several seconds behind the firing of neurons making the BOLD fMRI signal a coarse reflection of the neuronal activity. Despite this limitation, fMRI studies have increased geometrically over the last ten years, largely because the spatial resolution of fMRI is so high.

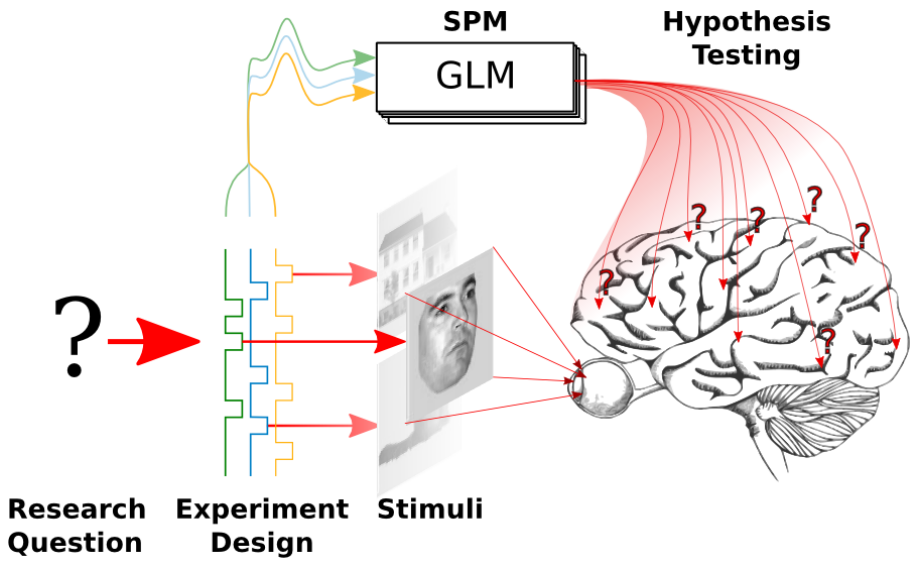
Researchers interested in the function of the human brain often focus on one, or at most a small subset of modalities. This is due partly because of the kinds of questions under investigation and partly to the heavy cost of learning to analyze data using other techniques. However, the selection of a given measurement approach can greatly limit the kind of research hypothesis that is formulated as well as the questions that can be addressed.

1.2 Machine Learning Methods in Neuroimaging

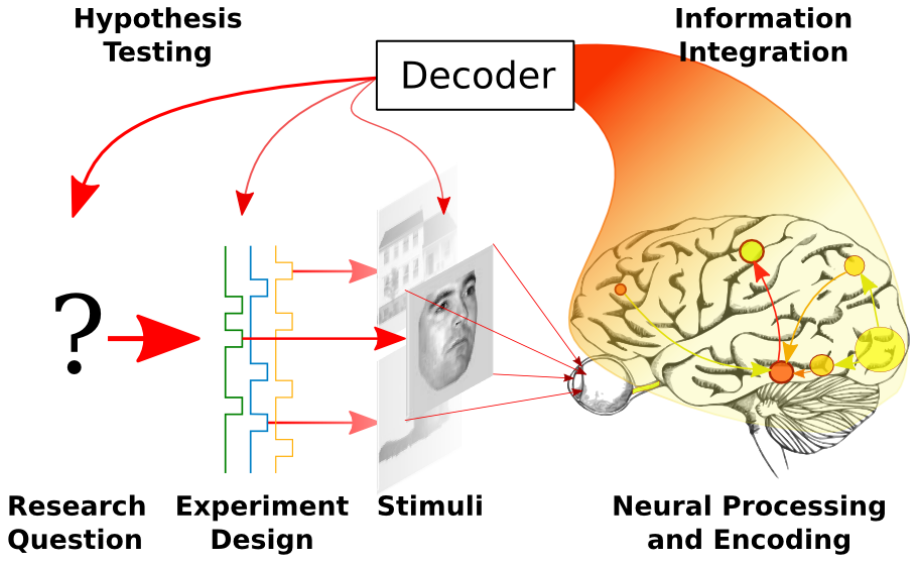
Some analysis techniques have become *de facto* standards despite their limitations or inappropriate assumptions for a given data type. For instance, the general linear model (GLM), as a part of statistical parametric mapping (SPM) analysis pipeline, is the prevalent approach used in fMRI data analysis. In SPM, aspects of neural signal processing were segregated into two independent steps: *identification* and *integration*. The original goal of *identification* was motivated by the intention to "identify functionally *specialized* brain responses" [12], *i.e.*, to identify regions of the brain (at the level of neural networks), where a statistically significant portion of the neural signal could be well described by known experimental manipulations (see Figure 1.4(a)). Such formulation of the target goal allowed significant simplification in the analysis of massive datasets, such as fMRI data, by considering each acquired spatial unit (*voxel*[5] in MRI terms) as independent of any other. This assumption promoted the use of *mass-univariate* analysis techniques. However, this approach is very restrictive in terms of possible findings [13, 14] because the analysis of each voxel is carried out in isolation from the rest of the population. Because of this fact, it impossible for a GLM analysis to account for any interaction within and between neural networks. Therefore the actual results of *identification* are limited to the regions with uniform contrast across experimental conditions, unless the interactions were explicitly modulated by the experimental design and were provided as input to the GLM. Due to such reliance upon large gross-effects within the signal, the GLM is not able to resolve the functional structures at the level of neural networks. By virtue of its design, the GLM is, at best, able to identify areas participating in the *encoding* of experimental variables into the neural code of the acquired neural data.

However, outside the neuroscience community, research in machine learning (ML) and statistical learning theory in particular, has spawned a set of analysis techniques that

[5] *Voxel* as a volume element, representing a value on a regular grid in three dimensional space.



(a) SPM via GLM



(b) Decoder

Figure 1.4 Two possible approaches toward the analysis of neural data (on the example of fMRI).

are typically multivariate, flexible (*e.g.*, classification, regression, clustering), powerful (*e.g.*, linear and non-linear), and generic, and hence often applicable to various data modalities with minor modality-specific pre-processing [see 15, for a tutorial and further reasoning on application of ML methods to the analysis of fMRI data]. Moreover, large parts of the ML community favor the open-source software development model [16, see also *MLOSS*[6] project website], which leads to an increase in scientific progress due to the superior accessibility of information and reproducibility of scientific results.

By virtue of advances in statistical learning techniques, various ML methods, such as support vector machines (SVM) [17], became capable of providing unprecedented generalization performance when tested on the data which were not provided during training. The ability to assess performance on data not seen during training provides the means of assessing the validity of a constructed decoder. Surprisingly, many methods retain a high level of generalization performance even when the dimensionality of the input space greatly exceeds the number of available data samples. Features such as these make statistical learning methods a perfect choice to tackle the problem of reliable *decoding* of neural data with its rich informational content and its multivariate nature.

Advantages of statistical learning methods have recently attracted considerable interest throughout the neuroscience community. Neuroscientists have reported compelling results when they applied such methods toward the analysis of fMRI data[7]. The major difference of such approaches from the "standard" SPM is the flow of information. In SPM, the description of the experimental conditions is provided as the input to GLM for a voxel by voxel testing of the desired hypothesis (Figure 1.4(a)). When using statistical learning methods, flow of information is usually reversed: entire (or a selected region of interest, or ROI) neural dataset is provided as the input to the

[6] http://www.mloss.org

[7] In the literature, authors have referred to the application of machine learning techniques to neural data as *decoding* [18, 19], *information-based* analysis [*e.g.*, 20] or *multi-voxel pattern analysis* [*e.g.*, 21, 22, 23].

decoder, which is trained to decode experimental conditions out of the observed neural data (Figure 1.4(b)). For example, two such classifier-based studies by Haynes and Rees [24] and Kamitani and Tong [18] were able to decode the orientation of visual stimuli from fMRI data from human primary visual cortex. Orientation-selective columns are only about 300–700 μm in width in monkeys, and on a millimeter scale in humans. For the reasons stated earlier, univariate methods are incapable of providing specificity information for a given spatial location at the level of neural networks. However, statistical learning methods can be used successfully with large data sets to reveal subtle response biases which would not be detected by univariate analysis Kamitani and Tong [18], Haynes and Rees [24]. These small signal biases representing local distributed coding of the responses are sufficient to disambiguate stimulus orientations despite the fact that the fMRI data were recorded at 3 mm spatial resolution, *i.e.*, at much more coarse resolution than the scale of the neural networks of interest. Thus, multivariate methods can be used to account not only for the variance in signal due to the interactions between voxels, but to extract information about functional structures from a signal acquired at a coarse level of spatial resolution.

Other classifier-based studies have further highlighted the strength of a *decoding* approach. For example, classifier-based analysis was first used to investigate neural representations of faces and objects in ventral temporal cortex and showed that the representations of different object categories are spatially distributed and overlapping and that they have a similarity structure that is related to stimulus properties and semantic relationships [25, 26, 27]. This dissertation work (Section 3.4.3, also includes work [28]) is examining the specificity information encoded in distributed patterns. The obtained results provide further understanding of the distributed similarity patterns obtained in the analysis of the data from [25].

Statistical learning methods have applications in the analysis of other neural data modalities as well. For instance, brain-computer interfaces (BCI) was an early research

direction that heavily relied on the use of ML methods. In the context of BCI, a trained classifier provides predictions about an intended action by a human based on the acquired neural data (usually from E/MEG modalities). Unlike research designed to uncover the functional structure of the brain, BCI work is concerned primarily with the detection of components of brain signals that a trained participant could easily control to perform a desired action (*i.e.*, moving a cursor around the screen). This dissertation research seeks not only to construct a decoder of neural data, but also to enhance understanding of the functional structure of the brain. Hence, Section 3.3.2 not only presents the ability to reliably predict experimental condition, but also shows how a constructed decoder of E/MEG signals could discover spatio-temporal components of the signal relevant to the research question, which are not detected by conventional univariate *identification* methods.

Besides BCI applications that use statistical learning algorithms in E/MEG, a variety of unsupervised statistical learning methods have recently received a lot of attention. Most of unsupervised methods address the problem of extracting "interesting" components of the signal by the means of the signal decomposition (*e.g.*, principal and independent component analysis (PCA and ICA)). These methods are used to decompose an acquired signal into a combination of basic components, where each component represents a specific neural or physiological process, or simply the instrumental noise. On the cellular level, such methods are usually used to decompose neuronal spike trends into a set of independent units that are believed to correspond to the basic elements of neural networks, *i.e.*, neurons. Since such *identification* of the neurons is driven by unsupervised methods, no *specificity* information is attributed to any given neuron in the result. Section 3.2 shows how supervised methods in the context of neural data *decoding* at the cellular level can provide specificity information for each extracted unit thereby making it possible to derive various conclusions about the neural population, signals of which are captured at the given recording site.

Unsupervised methods have also been used in pre-processing of data. The simplest application is the removal of the components which can be attributed to noise with some confidence thereby increasing the efficiency of data filtering. Alternatively such methods can be used for dimensionality reduction when only a few out of a number of components are found to be relevant to the experimental design. However, it is often the case that the components of interest are not the ones that account for most of the variance in the signal (see [13] or "Classification of SVD-mapped Datasets" in PyMVPA manual [29]).

To summarize, statistical learning methods have been used extensively in various fields of science and more recently in neuroscience. Nevertheless, only a subset of statistical learning methods are used in the analysis of neural data to discover the functional structure of the brain. Chapter 3 presents various strategies for applying ML methods to the analysis of various neural data modalities at different levels of investigation. It starts by addressing the problem of specificity at the neural level with the analysis of extracellular recordings. It then describes the analysis of E/MEG datasets and the advantages of *decoding* methods over standard univariate approaches in the identification of spatio-temporal components of the signal relevant to the experimental paradigm. The final section address once more the problem of *identification* and similarity structure analysis in fMRI data.

1.3 Multimodal Analysis

No single neural data modality has yet become the favored choice for resolving all neuroscience questions about brain function. High temporal resolution of E/MEG modalities is crucial in many event-related (ER) experiments but is absent in BOLD fMRI, a modality that delivers superior spatial resolution but poor temporal resolution. At this point in time a methodology that consolidates the information obtained from different brain imaging modalities and maximizes both spatial and temporal resolution

does not exist although such information integration would be extremely valuable in a variety of applications. It might provide a consistent and reliable localization of the functional neuronal units and processes with higher spatial and temporal precision that cannot be achieved using any of the existing modalities alone. It can help to understand the inherent structure of the signal: which parts of the observed signal could be attributed to the neural and physiological processes, and which to the instrumental noise, that is specific to a given neural modality in question. That can further lead to the development of cross-modality filtering methods, which could be used as a preprocessing step in the unimodal data analysis.

The main obstacle in the development of multimodal methods involving fMRI nowadays seems to be the absence of a reasonable universal model for hemodynamics, where the neuronal activation is the primary input factor. The naive models are not general enough to explain the variability of the BOLD signal, whereas complex parametric models that rely heavily on a prior knowledge of nuisance parameters (based on biophysical details), almost never have a reliable and straightforward means of estimation. This fact makes it unlikely to use such comprehensive models as reliable generative models of the BOLD signal at the level of basic neural networks.

Due to the difficulties in assessing the ground truth of a combined signal in any realistic experiment — a difficulty further confounded by lack of accurate biophysical models of BOLD signal, any cross-modality *fusion* approach has to be tackled with caution. Nevertheless, various methodologies can already provide basic test-ground to check the validity of the developed fusion methods: analysis of simplistic experiments, which target perception processes, where there is a small number of isolated focal sources of activity, studies on phantom (artificial) brains with known sources of activation. At last, the employment of the methods which enforce testing of the discovered underlying model on yet unseen data (which is a requisite in most of the

supervised ML methods), intrinsically provides unbiased estimation of the performance of the suggested method.

To summarize, now it seems to be the right time for the development of fusion methods which are comprising empirically supported models or are flexible enough to incorporate future elaborated models of the BOLD response. A convincing demonstration of obtaining better explanatory power for the signals and increased accuracy of localization by the means of multimodal integration for a complex protocol would constitute a major success in the field.

Chapter 4 addresses the development and validation of a multimodal functional brain imaging technique to gain intrinsic advantages of each used modality. Regression methods, devised in the field of statistical learning, are employed for the conjoint analysis of fMRI and EEG data, where fMRI data is explained in terms of neural activity registered by EEG modality. Such approach provides multiple advantages over existing methods which are presented in the same chapter. Proposed methodology is supported by the results obtained on simulated and empirical neural data.

1.4 Statistical Learning Software and Neuroimaging

Despite the aforementioned advantages and promises of statistical learning theory methods for the analysis of neural data, various factors have delayed their adoption for the analysis of neural information. First and foremost, existing conventional techniques are well-tested and often perfectly suitable for the standard analysis of data from the modality for which they were designed. Most importantly, however, a set of sophisticated software packages has evolved over time. That allows researchers to apply conventional and modality-specific methods without requiring in-depth knowledge about low-level programming languages or underlying numerical methods. In fact, most of these packages come with convenient graphical and command line interfaces. Presence

of such interfaces abstracts the peculiarities of the methods and allows researchers to focus on designing experiments and on addressing actual research questions without having to develop specialized analyses for each study.

While specialized software packages are useful when dealing with specific properties of a single data modality, they limit the flexibility to transfer newly developed analysis techniques to other fields of neuroscience. This issue is compounded by the closed-source, or restrictive licensing of many software packages. That, in fact, further limits software flexibility and extensibility.

Moreover, only a few software packages exist that are specifically tailored towards straightforward and interactive exploration of neuroscientific data using a broad range of ML techniques: Matlab®[8] MVPA toolbox for fMRI data[9][30] and *3dsvm* plugin for AFNI [31]. At present only independent component analysis (ICA), an unsupervised method, seems to be supported by numerous software packages (see 32, for fMRI, and 33, for EEG data analysis). Therefore, the application of statistical learning analyses strategies to neuroimaging data usually involves the development of significant amount of custom code. Hence, users are typically required to have in-depth knowledge about both data modality peculiarities and software implementation details.

Growing number of published results on using statistical learning methods in neuroimaging, coupled with the absent software implementation of those analysis pipelines, often makes it difficult, if not impossible, to carry out the same analysis on the datasets at hands to check the validity of the method, or barely to reproduce the reported results.

At the same time, Python has become the open-source scripting language of choice in the research community to prototype and carry out scientific data analyses and to develop complete software solutions quickly. It has attracted attention due to its openness,

[8]Closed source commercial product of MathWorks®
[9]It is possible to use the low-level functions of this toolbox for other modalities.

flexibility, and the availability of a constantly evolving set of tools for the analysis of many types of data. Python's automatic memory management, in conjunction with its powerful libraries for efficient computation (*NumPy*[10] and *SciPy*[11]) abstracts users from low-level "software engineering" tasks and allows them to fully concentrate their attention on the development of computational methods.

As an interpreted, high-level scripting language with a simple and consistent syntax, a plethora of available modules, easy ways to interface to low-level libraries written in other languages[12] and high-level computing environments[13], Python is the language of choice for solving many scientific computing problems. Despite the fact that it is possible to perform complex data analyses solely within Python, it *once again* often requires in-depth knowledge of numerous Python modules, as well as the development of a large amount of code to lay the foundation for one's work. Therefore, it would be of great value to have a framework that helps to abstract from both data modality specifics and the implementation details of a particular analysis method. Ideally, such a framework should help to expose any form of data in an optimal format applicable to a broad range of machine-learning methods, and on the other hand provide a versatile, yet simple, interface to plug in additional algorithms operating on the data.

Chapter 5 presents PyMVPA — a Python-based framework for multivariate pattern analysis using supervised and unsupervised ML methods. The chapter provides a short summary of the principal design concepts, and the basic building blocks of the PyMVPA. This software was used for most of the the analysis results presented in this dissertation, and major parts of the source code for the analysis are readily available under FOSS license as part of PyMVPA distribution, or as supplemental material for [23].

---

[10] http://numpy.scipy.org
[11] http://www.scipy.org
[12] *e.g.*, ctypes, SWIG, SIP, Cython
[13] *e.g.*, mlabwrap and RPy

CHAPTER 2

RESEARCH METHODOLOGY

This works every time, provided you're lucky

– Unknown soul-mate

2.1 Machine Learning

Machine learning methods aim at integrating the available knowledge extracted from the provided data, so that they could reliably assess the outcome whenever new data is provided. Significant boost in the applicability of machine learning toward the analysis of real data came with the development of statistical learning theory. Moreover, all of the methods utilized in this dissertation were developed in the domain of statistical learning. Nevertheless, "machine learning" is used throughout the text as a generic term for any methodology which targets to "train" a reliable predictor of a novel data.

Machine learning (ML) methods employed in this dissertation could be brought into two groups: supervised and unsupervised. A supervised learning method requires labeled data, so it could deduce the mapping from the data onto the labels. This step is called *learning*. The quality of the learned mapping is afterward assessed by applying it to new data and comparing mapping results to the true labels. This step is called *generalization testing* or simply *testing*. Unsupervised methods do not require the labeling of the data provided for learning. Usually these methods are constructed to discover an inherent structure of the input data (*e.g.*, maximum variance linear subspaces in PCA, independent components in ICA, *etc.*).

There exists a multitude of ML methods. They differ in their goals, assumptions about the underlying data structure, inherent feature weighting and selection, target

optimization criterions, optimization routines, and many other aspects. Reliable decoding of the neural data, which is the topic of this dissertation, does not aim at exploring and describing possible learning methods in the search for the best choice. Rather, it concerns itself with the construction of a generic workflow, methods, and tools capable of extracting information out of the neural data at different levels of investigation of human brain.

ML-based analysis typically consists of some basic procedures that are independent of the classification algorithm or decision process that was actually used, *e.g.*, error calculation, cross-validation of prediction performance, and feature selection. Some of the aspects of learning methods presented in this chapter are presented in the perspective of PyMVPA framework, which is described in greater detail in the Chapter 5. Meanwhile, Figure 2.1 presents building blocks of PyMVPA to highlight various stages of data analysis.

2.2 Data Handling

Typically, a dataset for ML analysis consists of three parts: *data samples*, *sample attributes* and *dataset attributes*. While the data samples contain the actual patterns that shall be used for training or validation, sample attributes hold additional information on a per sample basis (see Figure 2.2). First and foremost of these are the discrete labels (classification) or continuous variables (regression) that index each data sample with a given experimental condition and thereby provide the basis for the mapping that is constructed by the learning method.

It is often necessary to define groups of data samples. For example, when performing cross-validation, the independent training and validation sets must be identified. In the case of fMRI data, which contains significant forward temporal contamination across samples, the temporal separation of training and validation datasets

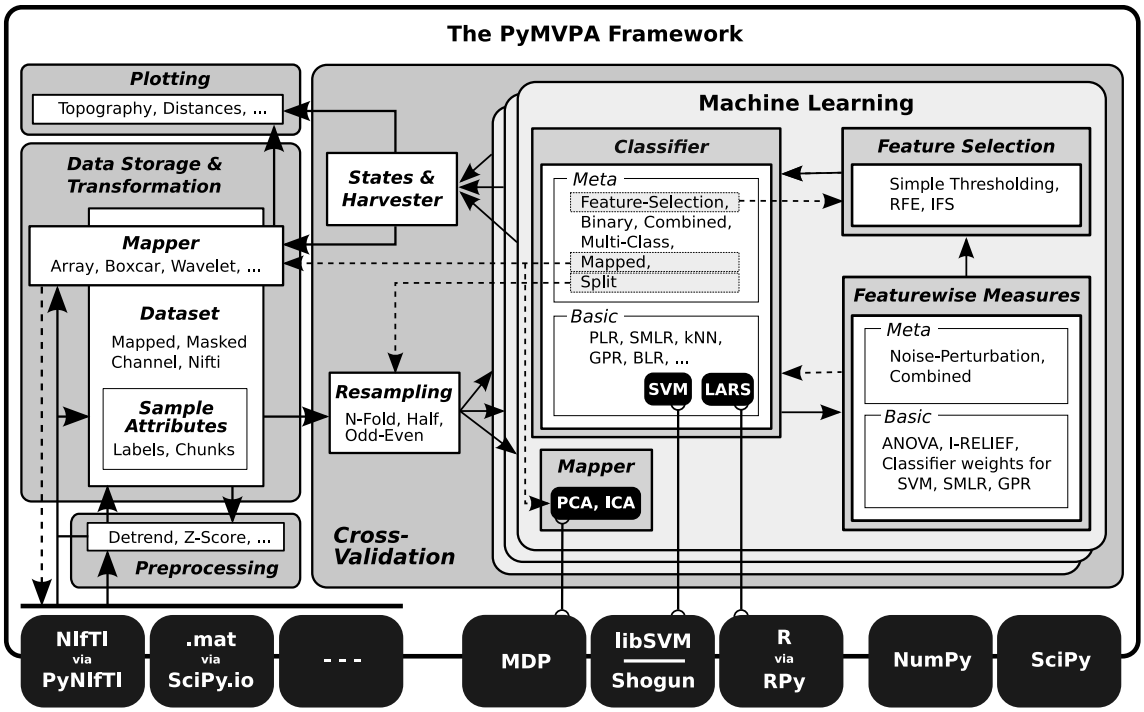


Figure 2.1 Machine learning framework PyMVPA, its workflow and design. PyMVPA consists of several components (gray boxes) such as ML algorithms or dataset storage facilities. Each component contains one or more modules (white boxes) providing a certain functionality, *e.g.*, *classifiers*, but also feature-wise measures [*e.g.*, I-RELIEF; 34], and feature selection methods [recursive feature elimination, RFE; 35, 36]. Typically, all implementations within a module are accessible through a uniform interface and can therefore be used interchangeably, *i.e.*, any algorithm using a classifier can be used with any available classifier implementation, such as *support vector machine* [SVM; 17], or *sparse multinomial logistic regression* [SMLR; 37]. Some ML modules provide generic *meta* algorithms that can be combined with the *basic* implementations of ML algorithms. For example, a *Multi-Class* meta classifier provides support for multi-class problems, even if an underlying classifier is only capable to deal with binary problems. Additionally, most of the components in PyMVPA make use of some functionality provided by external software packages (black boxes). *SVM*, classifiers are interfaced to implementation in *Shogun* or *LIBSVM*. PyMVPA provides a convenient wrapper for access through a uniform interface. PyMVPA provides a simple yet flexible interface that is specifically designed to make use of externally developed software. An analysis built from these basic elements can be cross-validated by running them on multiple dataset splits that are generated through various data resampling procedures [*e.g.*, bootstrapping, 38]. Detailed information about the results of any analysis can be queried from any building block and can be visualized through various plotting functions supported by PyMVPA. Alternatively, the results can be mapped back into the original data space and format for further processing by specialized tools. The solid arrows represent a typical connection pattern between the modules. Dashed arrows refer to additional compatible interfaces which, although potentially useful, are not necessarily used in a standard processing workflow.

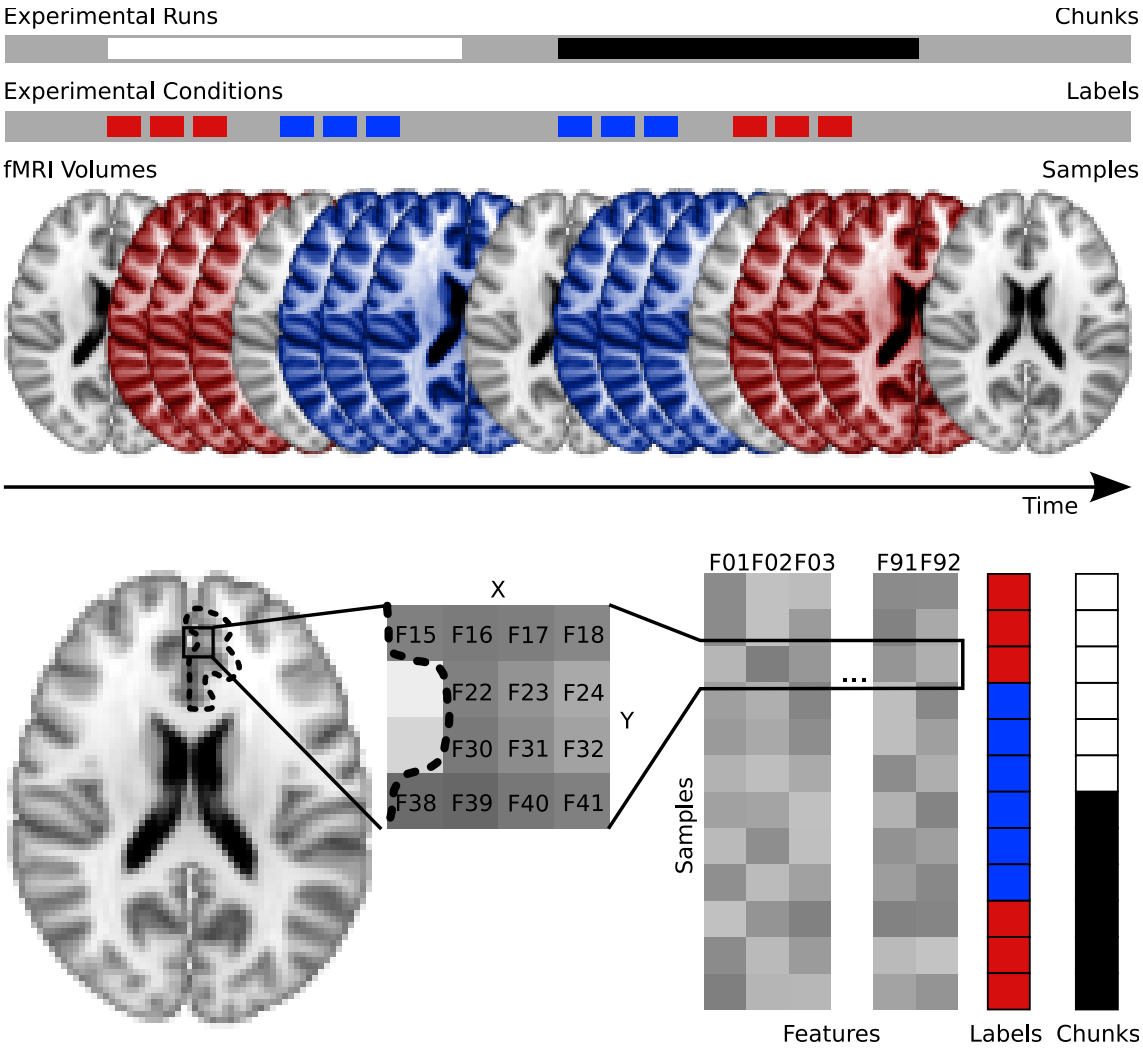


Figure 2.2 Terminology for ML-based analyses of fMRI data. The upper part shows a simple block-design experiment with two experimental conditions (*red* and *blue*) and two experimental runs (*black* and *white*). Experimental runs are referred to as independent *chunks* of data and fMRI volumes recorded in certain experimental conditions are data *samples* with the corresponding condition *labels* attached to them (for the purpose of visualization the axial slices are taken from the MNI152 template downsampled to 3 mm isotopic resolution). The lower part shows an example ROI analysis of that paradigm. All voxels in the defined ROI are considered as *features*. The three-dimensional data samples are transformed into a two-dimensional *samples*×*feature* representation, where each row (sample) of the data matrix is associated with a certain label and data chunk.

is mandatory. This is typically achieved by splitting an experiment into several runs that are recorded separately. In PyMVPA this type of information can be specified by a special *chunks* sample attribute, where each sample is associated with the numerical identifier of its data chunk or run (see Figure 2.2). It is possible to create any number of auxiliary sample attributes in a similar fashion.

Most machine learning software requires data to be represented in a simple two-dimensional $samples \times features$ matrix (see Figure 2.2, bottom), however, fMRI datasets are typically four-dimensional. Although it is possible to view each volume as a simple vector of voxels, doing so ignores information about the spatial properties of the volume samples. This is a particularly serious disadvantage given that spatial metrics, and especially distance information, are of interest in brain research. In addition, some analysis algorithms, such as the *multivariate searchlight* [20], make use of this information when calculating spheres of voxels.

PyMVPA follows a different approach. Each dataset is accompanied by a transformation or mapping algorithm that preserves all the required information and stores it as a dataset attribute. These mappers allow for bidirectional transformations from the original data space into the generic 2-D matrix representation and vice versa. In the case of fMRI volumes, the mapper indexes each feature with its original coordinates in the volume. It can optionally compute customizable distances (*e.g.*, Cartesian) between features by taking both voxel size and coordinates into account. Using the mapper in the reverse direction, from generic feature space into original data space makes it easy to visualize the results of the analysis. For example, feature sensitivity maps can be easily projected back into a 3-D volume and visualized in a way that is similar to a statistical parametric map.

2.3 Generalization Testing

During a typical classifier-based analysis, a particular dataset has to be resampled several times to obtain an unbiased generalization estimate of a specific classifier. In the simplest case, resampling is done by splitting the dataset so that some part is used for validation while the remainder is used for training. This is done multiple times until a stable estimate is achieved or the particular sampling procedure exhausts all possible ways of partitioning of the data. While the proper splitting of a dataset is very important, it might not be obvious how to do it given the forward temporal contamination across trials that is produced by the hemodynamic response function. Violating the strict separation of training and validation datasets will produce bias in subsequent analyses to the extent that the classifier has access to the data against which it will be validated.

To address this problem, PyMVPA provides a number of resampling algorithms. The most generic one is an N-M splitter where M out of N dataset chunks are chosen as the validation dataset while all others serve as training data until all possible combinations of M chunks are drawn. This implementation can be used for leave-one-out cross-validation, but also provides functionality useful for bootstrapping procedures [38]. Additional splitters are available that produce first-half-second-half or odd-even splits. Because any splitter may base its splitting on any given sample attribute, it is possible to split datasets based on data chunks or by *e.g.*, stimulus conditions.

Most algorithms implemented in PyMVPA can be parameterized with a splitter, making them easy to apply within different kinds of splitting or cross-validation procedures. As with other functions available in PyMVPA, the common interface makes it trivial to add custom splitters.

The dataset resampling functionality in PyMVPA also eases non-parametric testing of classification and generalization performance via a data randomization approach, *e.g.*, Monte Carlo permutation testing [39]. By running the same analysis multiple times with permuted dataset labels (independently within each data chunk) it is

possible to obtain an estimate of the baseline or chance performance of a classifier or some other sensitivity measure. This allows an estimation of statistical significance (in terms of p-values) for results achieved using the original (non-permuted) dataset.

2.4 Feature Sensitivity Analysis

A primary goal for brain-mapping research is to determine where in the brain information is processed and stored. Univariate analysis procedures typically provide localization information because each feature is tested independently of all others. In contrast, classifier-based analyses incorporate information from the entire feature set in order to determine whether or not the classifier can extract sufficient information to predict the experimental condition from the recorded brain activity. Although classifiers can use the joint signal of the whole feature set to make predictions, it is still important to know which features contribute to those correct predictions. Some classifiers readily provide information about *sensitivities*, *i.e.*, feature-wise scores measuring the impact of each feature on the decision made by the classifier. For example, a simple artificial neural network or a logistic regression, such as SMLR, bases its decisions on a weighted sum of the inputs. Similar weights can also be extracted from any linear classifier including SVMs. It is possible to compare the weights of the features to each other, if input data was standardized (z-scored) per each input feature to remove possible variance factor which invalidate such a direct comparison.

However, there are also *classifier-independent algorithms* to compute featurewise measures. While neural network and SVM weights are inherently multivariate, a feature-wise ANOVA, *i.e.*, the fraction of within-class and across class variances, is a univariate measure, as is simple variance or entropy measures of each voxel over all classes. In addition to a simple ANOVA measure, PyMVPA provides linear SVM, GPR, and SMLR weights as basic feature sensitivities. As with the classifiers discussed in the previous

section, a simple and intuitive interface makes it easy to extend PyMVPA with custom measures (*e.g.*, information entropy). The SciPy package, for example, provides a large variety of measures that can be easily used within the PyMVPA framework.

PyMVPA provides some algorithms that can be used on top of the basic featurewise measures to potentially increase their reliability. Multiple feature measures can be easily computed for sub-splits of the training data and combined into a single featurewise measure by averaging, t-scoring or rank-averaging across all splits. This is useful for stabilizing measure estimates, for example, when a dataset contains spatially distributed artifacts. While a GLM is rather insensitive to such artifacts because it looks at each voxel individually [40], classifiers are often able to pick up such a signal if it is related to the classification decision. Moreover, if artifacts are not equally distributed across the entire experiment, then computing measures for separate sub-splits of the dataset can be used to identify and reduce their impact on the final measure.

In addition to comparing raw sensitivity values, PyMVPA enables researchers to easily conduct *noise perturbation* analyses, where one measure of interest, such as cross-validated classifier performance, is computed many times with a certain amount of noise added to each feature in turn. Feature sensitivity is then expressed in terms of the difference between computed measures with and without noise added to a feature [see 26, 41, for equivalence analyses between noise perturbation and simple sensitivities for SVM].

2.5 Feature Selection Procedures

As mentioned above, featurewise measure maps can be easily computed with a variety of algorithms. However, such maps alone do not provide enough information to determine the features that are necessary or sufficient for performing some classification. However, feature selection algorithms can be used to determine the relevant features based on a

featurewise measure. As such, feature selection can be performed in a data-driven or classifier-driven fashion. In a data-driven selection, features can be chosen according to some criterion such as a statistically significant ANOVA score for the feature given a particular dataset, or a statistically significant t-score of one particular weight across splits. Classifier-driven approaches usually involve a sequence of training and validation actions to determine the feature set which is optimal with respect to some classification error (*e.g.*, transfer, inherent leave-one-out, theoretical upper-bound) value. It should be noted that to perform unbiased feature selection using the classifier-driven approach, the selection has to be carried out without observing the main validation dataset for the classifier.

Among the existing algorithms *incremental feature search* (IFS) and *recursive feature elimination* [RFE, 35, 36] are widely used [*e.g.*, 41] and both available within PyMVPA. These methods differ in what they use as a starting point and how they perform feature selection. RFE starts with the full feature set and attempts to remove the least-important features until a stopping criterion is reached. IFS on the other hand starts with an empty feature set and sequentially adds important features until a stopping criterion is reached. The implementations of both algorithms in PyMVPA are very flexible as they can be parameterized with all available basic and meta featurewise measures and the specifics of the iteration process and the stopping criteria are both fully customizable.

Results presented in this dissertation relied on RFE to improve generalization performance of SVM classifiers (Section 3.4.3). In other analyses, SMLR classifier with its automatic feature selection found to perform better, hence was used as the first preferable approach whenever dimensionality of the data was relatively large.

CHAPTER 3

UNIMODAL ANALYSIS OF NEUROIMAGING DATA

Воистину есть ложь, беспардонная ложь и статистика.
но не будем, друзья, забывать и о психологии!

– А. и Б. Стругацкие *"Жук в муравейнике"*, 1979

Truly, there are lies, brazen lies, and statistics, but let's not,
my friends, forget the psychology!

– A. and B. Stroogatskie *"The bug in an ant hill"*, 1979

3.1 Introduction

This chapter explores the application of statistical learning methods to the analysis of neural data at different levels of investigation. It begins with an analysis of multiple extracellular recordings (Section 3.2), where supervised learning methods are used to confirm that a given recording site contains neurons specialized for the given task. Inspection of the sensitivities of the trained classifiers further enriches analysis results. Classifier sensitivity analysis for each extracted neuron's temporal signal evolution allows to characterize specificity of that neuron in respect to a given stimuli condition. Temporally rich signals, such as E/MEG, which reflect neural activity of the whole brain, are analyzed next in Sections 3.3.2 and 3.3.3. Here, the suggested *decoding* approach makes it possible to unravel spatio-temporal components of the signal. The components do not show significant task-dependent change if analyzed by conventional univariate methods. Section 3.4 concludes the chapter with an extensive analysis of an fMRI dataset which carries spatially rich information about mechanisms of the brain involved in visual object processing.

Most of the analysis presented in this chapter was carried out within the PyMVPA framework. The comprehensive review of the PyMVPA framework is presented in the Chapter 5.

3.2 Extracellular Recordings

Because the extracellular dataset that will be analyzed in this section is unpublished, the experimental and acquisition parameters are described here for completeness and comprehensibility. Animal experiments were carried out in accordance with the National Institute of Health Guide for the Care and Use of Laboratory Animals and approved by Rutgers University. Sprague-Dawley rats (300-500 g) were anaesthetized with urethane (1.5 g/kg) and held with a custom naso-orbital restraint. After preparing a 3 mm square window in the skull over the auditory cortex, the dura was removed, and a silicon microelectrode consisting of eight four-site recording shanks (NeuroNexus Technologies, Ann Arbor, MI) was inserted. The recording sites were in the primary auditory cortex, estimated by stereotaxic coordinates, vascular structure [42] and tonotopic variation of frequency tuning across recording shanks, and located within layer V, as determined by electrode depth and firing patterns.

Five pure tones (3, 7, 12, 20, 30 kHz at 60 dB) and five different natural sounds (extracted from the CD “Voices of the Swamp”, Naturesound Studio, Ithaca, NY) were used as stimuli. Each stimulus had a duration of 500 ms followed by 1500 ms of silence. All stimuli were tapered at the beginning and end with a 5 ms cosine window. The data acquisition took place in a single-walled sound isolation chamber (IAC, Bronx, NY) with sounds presented free field (RP2/ES1, Tucker-Davis, Alachua, FL).

Individual units[1] were isolated by a semiautomatic algorithm (*KlustaKwik*[2]) followed by manual clustering (*Klusters*[3]). Post-stimulus time histograms (PSTH) of spike counts per each unit for all 1734 stimulus onsets were estimated using a bin size of 3.2 ms. To ensure the accurate estimation of PSTHs, only units with a mean firing rate higher than 2 Hz were selected for further analysis, leaving a total of 105 units for the analysis.

Because the segregation of recordings from individual units is performed independent of the associated stimuli, *i.e.*, in an unsupervised fashion (ML terminology), the activity of any particular unit can not be easily attributed to a particular stimulus. As can be seen in the top plots of Figure 3.2, the stimulus-wise descriptive statistics of the units presented make it difficult to pair the activity of a given unit at a particular time to a given stimulus. Furthermore, due to the inter-trial variance in the spike counts, it is even more difficult to reliably determine to what stimulus condition a given trial belongs. To eliminate the problems associated with unsupervised learning, an anlysis based on a reliable decoding of neural data, obtained from this limited neural population, was performed. This new analysis allowed the characterization of all extracted units in terms of their specificity to a given stimulus at a given point in time.

The analysis involved a standard 8-fold cross-validation procedure run with an SMLR [SMLR; 37] classifier, which achieved a mean estimated accuracy of 77.57% across all 10 types of stimuli. This generalization accuracy was well above chance (10%) for all stimulus categories and provided evidence for a differential signal across the 10 stimuli at the recording site. Misclassifications were found primarily for low-frequency stimuli. Pure tones with 3 kHz and 7 kHz were more often confused with each other than were tones with a larger frequency difference (see Figure 3.1), suggesting a high similarity

[1]The term "unit" here refers to a single entity, which was segregated from the recorded data, and is expected to represent a single neuron.
[2]http://klustakwik.sourceforge.net
[3]http://klusters.sourceforge.net

in the spiking patterns for these frequencies. Moreover, the greater discriminability for stimuli at higher frequencies suggests that this neuronal population may be specialized for processing of higher frequency tones.

In addition to labelling unseen trials with high accuracy, the trained classifier provided sensitivity estimates for each unit, time bin, and stimulus condition (see bottom plots of Figure 3.2). Stimulus specific information that is contained in the spike times, relative to a given stimulus onset, can be determined in the temporal sensitivity profiles of any particular unit (see unit #42 profiles in lower left plot of Figure 3.2). Alternatively, stimulus specificity could be represented as a slowly modulated pattern of spike counts (see 3 kHz stimuli). The aggregate sensitivity (in this case the sum of absolute sensitivities) across all time-bins provides summary statistics for any unit's sensitivity to a given stimulus (see lower right plot of Figure 3.2), and unlike a simple variance measure, the aggregate sensitivity makes it relatively easy to associate any unit to any set of stimuli. An additional advantage provided by this approach is the ability to identify units whose signals do not display a substantial amount of variance, but nevertheless carry a stimulus specific signal (*e.g.*, unit #28 and 30 kHz stimulus).

3.3 EEG and MEG

3.3.1 Introduction

Conventional analysis approaches of E/MEG often rely on averaging over multiple trials to extract statistically relevant differences between two or more experimental conditions (*e.g.*, averaged ERP [43]). This practice of averaging remains dominant in the investigation of neural processes and relies on an assumption that response does not vary across trials of the same experimental condition. In the last decade, though, statistical decomposition methods (*e.g.*, ICA) have been used more frequently in the analysis of E/MEG data. The popularity of these methods is increasing because they provide a way

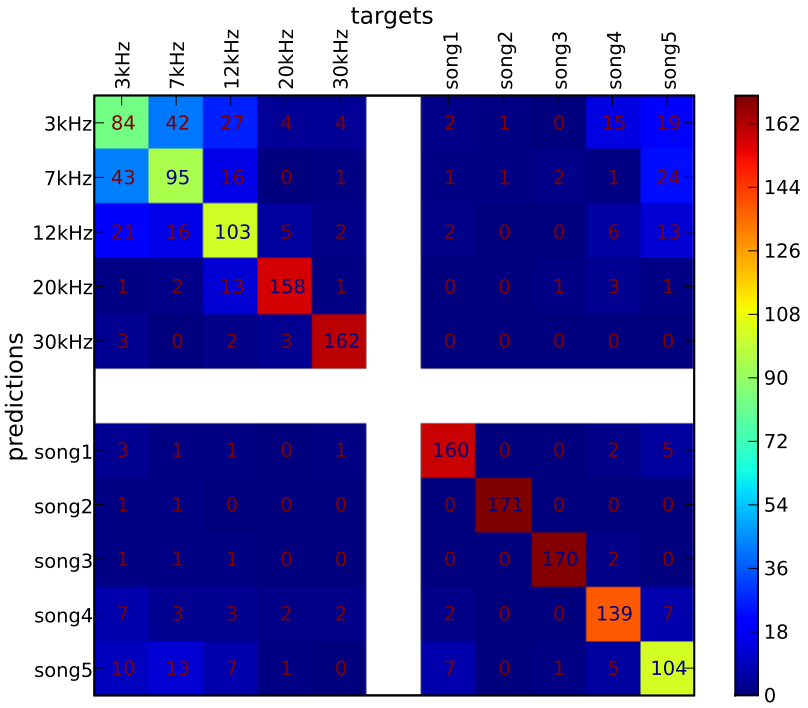


Figure 3.1 Confusion matrix of SMLR classifier predictions of stimulus from multiple single units recorded simultaneously. The classifier was trained to discriminate between stimuli of five pure tones and five natural sounds. Elements of the matrix (numeric values and color-mapped visualization) show the number of trials which were correctly (diagonal) or incorrectly (off-diagonal) classified by a SMLR classifier during an 8-fold cross-validation procedure. The results suggest a high similarity in the spiking patterns for stimuli of low-frequency pure tones, which lead the classifier to confuse them more often, whereas responses to natural sound stimuli and high-frequency tones were rarely confused with one another.

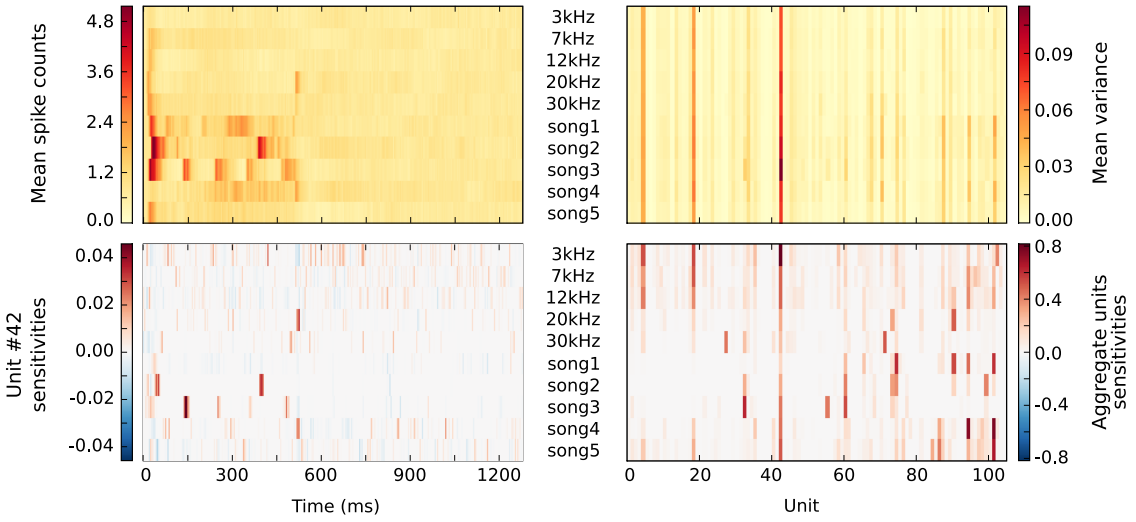


Figure 3.2 Statistics of multiple single unit extracellular simultaneous recordings and corresponding classifier sensitivities. All plots sweep through different stimuli along vertical axis, with stimuli labels presented in the middle of the plots. The upper part shows basic descriptive statistics of spike counts for each stimulus per each time bin (on the left) and per each unit (on the right). Such statistics seem to lack stimulus specificity for any given category at a given time point or unit. The lower part on the left shows the temporal sensitivity profile of a representative unit for each stimulus. It shows that stimulus specific information in the response can be coded primarily temporally (few specific offsets with maximal sensitivity like for *song2* stimulus) or in a slowly modulated pattern of spikes counts (see 3kHz stimulus). Associated aggregate sensitivities of all units for all stimuli in the lower right figure indicate each unit's specificity to any given stimulus. It provides better specificity than a simple statistical measure such as variance, *e.g.*, unit 19 active in all stimulation conditions according to its high variance, but according to the classifier sensitivity it carries little, if any, stimuli specific information for natural songs 1-3.

to filter out irrelevant variance contributions and to choose only the components which may be relevant to the experimental design. This feature makes it possible to carry out analysis of E/MEG data on a per trial basis [44, 45]. Per trial analysis of E/MEG data offers new insight into the associated underlying processes of the data by addressing the variability of neural activation across trials [46, 47].

The decoding of E/MEG signals with supervised learning methods has been used most extensively in the construction of the brain-computer interfaces. Trained classifiers are often capable of detecting a targeted action type and of predicting a subject's goal for that action. The bulk of this work, therefore, is focused on practical application rather than scientific discovery. However, supervised learning methods can be used to unlock some of the mysteries of the brain. For example, one early study [48] optimized a "spatial" linear discriminator (using a predefined time window during training) which provided localization of the corresponding brain area. Even so, such approaches have not been geared toward automatic detection of both spatial and temporal components of the signal relevant to the experimental design.

This next section explores the application of statistical learning methods for the per-trial analysis of E/MEG data. It will be demonstrated that sensitivity analysis of the classifiers reveals spatio-temporal components which are not detected using conventional methods such as ERP.

3.3.2 EEG

The dataset used for the EEG example consists of a single participant from a previously published study on object recognition [49]. In the experiment, participants indicated, for a sequence of images, whether they considered each particular image a meaningful object or just "object-like." This task was performed under two different speed constraints for two sets of stimuli having different statistical properties. EEG was recorded from 31

electrodes at a sampling rate of 500 Hz using standard recording techniques. Details of the recording procedure can be found in Fründ et al. [49]. A detailed description of the stimuli can be found in Busch et al. [50, colored images] and in Herrmann et al. [51, line-art pictures].

Fründ et al. [49] performed a wavelet-based time-frequency analyses of channels from a posterior region of interest (*i.e.*, no multivariate methods were employed). However, here multivariate methods are used to differentiate between trials with colored stimuli (having a broad spectrum of spatial frequencies and a high level of detail) and trials with black and white line-art stimuli (Figure 3.3A). This discrimination is orthogonal to the experimental task that required participants to distinguish between object and non-object stimuli.

The data for this analysis were 700 ms EEG segments starting 200 ms prior to the stimulus onset of each trial. Only trials that passed the semi-automatic artifact rejection procedure performed in the original study were included, yielding 852 trials (422 color and 430 line-art). Each trial timeseries was downsampled to 200 Hz, leaving 140 sample points per trial and electrode. Then each trial was defined by including the EEG signal of all sample points from all channels, as a sample to be classified (4340 features total). Finally, all features for each sample were normalized to a zero mean and unit variance (z-scored).

As the main analysis a standard 6-fold cross-validation procedure was applied to *linear support vector machine* [linCSVM; 17], *sparse multinomial logistic regression* [SMLR; 37] and *Gaussian process regression* with linear kernel [linGPR; 52] classifiers. Additionally, the multivariate I-RELIEF [34] feature sensitivity measures, and, for comparison, a univariate analysis of variance (ANOVA) F-score, were computed on the same cross-validation dataset splits.

All three classifiers performed with high accuracy on the independent test datasets, achieving 86.2% (linCSVM), 91.8% (SMLR), and 89.6% (linGPR) correct

single trial predictions, respectively. However, of greater interest are the features each classifier used for predictions. Using PyMVPA it is very easy to extract feature sensitivity information from all used classifiers. Figure 3.3B shows the computed sensitivities from all classifiers and measures. There is a striking similarity between the shape of the classifier sensitivities plotted over time and the corresponding event-related potential (ERP) difference wave between the two experimental conditions (Figure 3.3A; example shown for electrode *Pz*, Fründ et al. [49]). Total (mean across all time-points) head topography plots of the sensitivities (left column of Figure 3.3C) reveals a high variability with respect to the specificity among the multivariate measures. SVM, GPR and SMLR weights congruently identify three posterior electrodes as being most informative (SMLR weights provide the highest contrast of all measures). The I-RELIEF topography is much less specific and more similar to the ANOVA topography in its global spatial structure than to the other multivariate measures. It should be noted, however, that these topographies aggregate information over all timepoints and, therefore, do not provide information about specific temporal EEG components, which are presented in the corresponding columns of Figure 3.3C.

Of particular interest is the comparison between the multivariate sensitivities and the univariate ANOVA F-scores from 300 ms to 400 ms following stimulus onset (right most column of Figure 3.3C demonstrates the sensitivities topographies at 370 ms). Only the multivariate methods (especially SMLR, linCSVM and linGPR) detected a relevant contribution to the classification task of the signal in this time window. This late signal may be related to intracranial EEG gamma-band responses that Lachaux et al. [53] are observed at around the same time that participants viewed complex stimuli. Given that the present data also seem to show a similar evoked gamma-band response [49], it is possible that multivariate methods are sensitive to gamma-band activity in neural data.

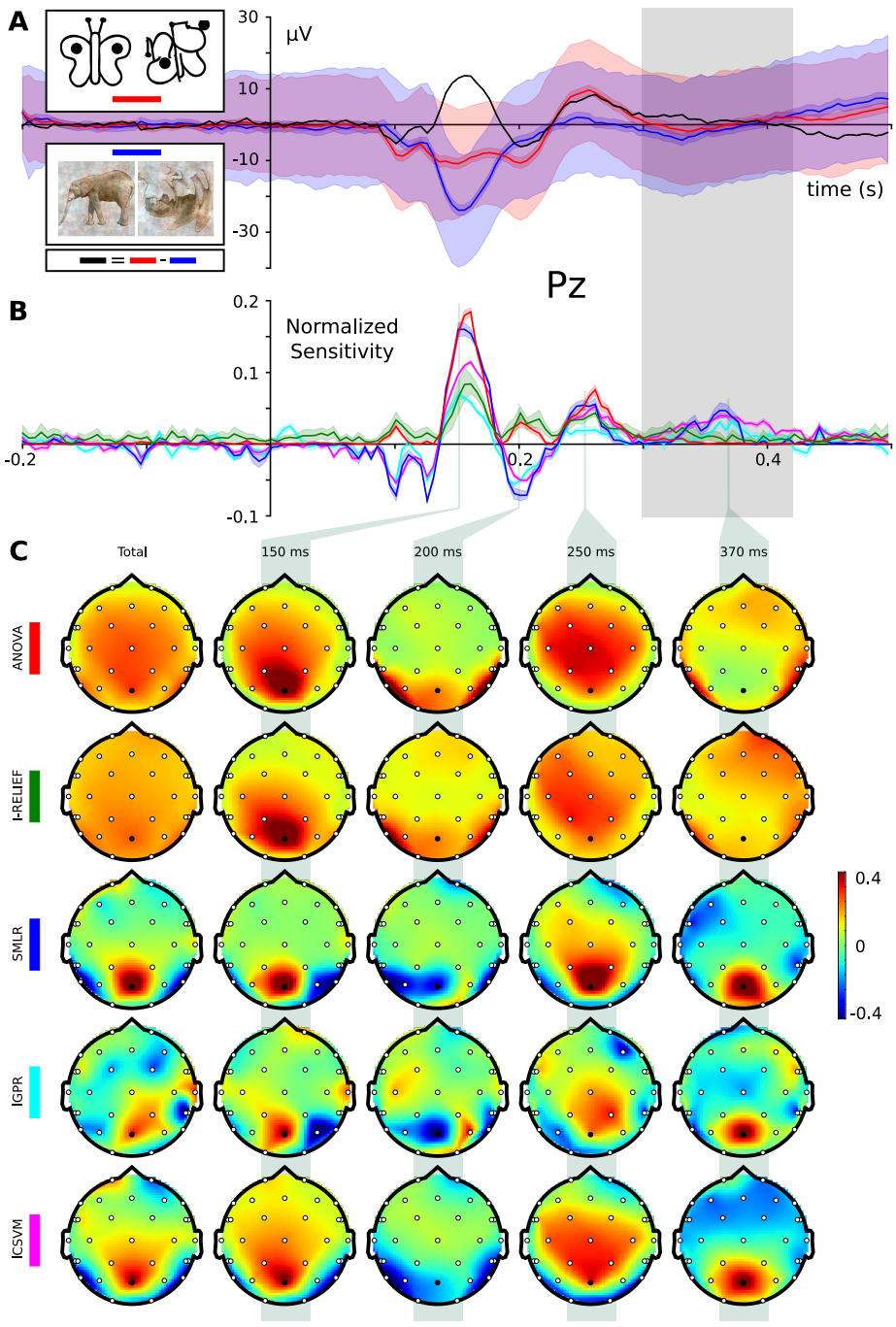


Figure 3.3 Event-related potentials and classifiers sensitivities for the classification of color and line-art conditions in EEG data.

3.3.3 MEG

The example MEG dataset [54] was collected with the goal of determining whether it is possible to predict the recognition of briefly presented natural scenes from single trial MEG-recordings of brain activity using statistical learning methods. On each trial participants saw a briefly presented photograph (37 ms) depicting a natural scene followed immediately by a pattern mask (1000 ms – 1400 ms). The short interval between presentation and mask effectively contrains the amount of processing that can be done [55] and, therefore, limits the number of pictures that participants will later recognize. After the mask was turned off, participants predicted (using a button press response) whether or not they would be able to recognize the picture that had just been presented. Immediately after this prediction, partipants were presented with four natural scene photographs and asked to indicate which of the four scenes had been presented (*i.e.*, a four-alternative forced-choice delayed match to sample task).

The MEG was recorded with a 151 channel CTF Omega MEG system from the whole head (sampling rate 625 Hz and a 120 Hz analogue low pass filter) while participants performed this task. The 600 ms interval of the MEG time series data that was used for the analysis started at the onset of the briefly presented scene and ended before the mask was turned off. As in the original study, only trials in which predictions matched recognition performance (*i.e.*, *RECOG-RECOG*, *NRECOG-NRECOG*) were analyzed. For details about the rationale of this selection, the stimulus presentation information, and the recording procedure see Rieger et al. [54]. In this analysis data from a single participant (labeled *P1* in the original publication) was used.

The MEG timeseries were first downsampled to 80 Hz and then all trial segments were channel-wise normalized by subtracting their mean baseline signal (determined from a 200 ms window prior to scene onset). Only timepoints within the first 600 ms after stimulus onset were considered for further analysis. The resulting dataset consisted of

151 channels with 48 timepoints each (7248 features), and a total of 294 samples (233 *RECOG* trials and 61 *NRECOG* trials).

The original study contained analyses based on SVM classifiers, which revealed, by means of the spatio-temporal distribution of the sensitivities, that the theta band by itself provides the most discriminative signal. The authors also addressed the issue of interpreting heavily unbalanced datasets[4]. Given this comprehensive analysis, the goal of the analysis presented here was to determine if this analysis strategy could be replicated within PyMVPA.

As with the EEG data, a standard cross-validation procedure was applied (this time 8-fold), using linear SVM and SMLR classifiers. Additionally, again univariate ANOVA F-scores were computed on the same cross-validation dataset splits. The SVM classifier was configured to use different per-class C-values[5], scaled with respect to the number of samples in each class to address the unbalanced number of samples, which was not done by the original authors. Similar to Rieger et al. [54], a second cross-validation on balanced datasets (by performing multiple selections of a random subset of samples from the larger *RECOG* category) was run.

Both, classifiers performed almost identically on the full, unbalanced dataset, achieving 84.69% (SMLR) and 82.31% (linCSVM) correct single trial predictions (83.0% in the original study). Figure 3.4 shows sample timeseries of the classifier sensitivities and the ANOVA F-score of two posterior channels. Due to the significant difference in the number of samples of each category, it is important to note that the mean true positive rate (TPR)[6], that amounted to 72% (SMLR), and 76% (linCSVM) respectively. The second

[4]Unbalanced datasets have a dominant category which has considerably more samples than any other category. That potentially leads to the problem when a classifier prefers to assign the label of that category to all samples to minimize total prediction error.

[5]Parameter C in soft-margin SVM controls a trade-off between width of the SVM margin and number of support vectors [see 56, for an evaluation of this approach]

[6]Mean TPR is equal to the accuracy in balanced sets, and is at 50% chance performance even with unbalanced set sizes [see 54, for a discussion of this point].

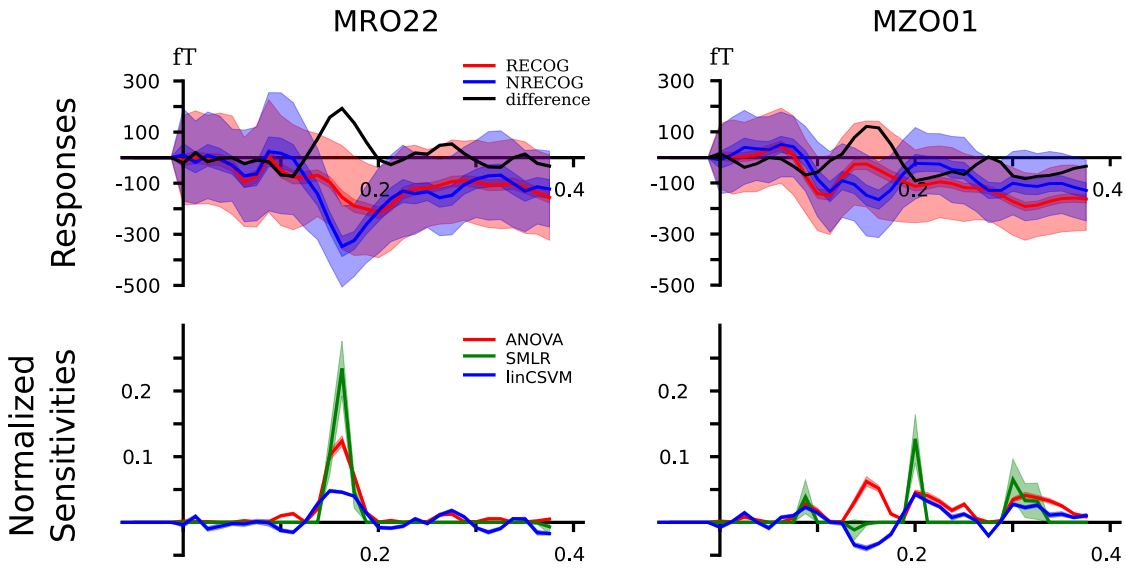


Figure 3.4 Event-related magnetic fields and classifier sensitivities. The upper part shows magnetic fields for two exemplary MEG channels. On the left sensor MRO22 (right occipital), and on the right sensor MZO01 (central occipital). The lower part shows classifier sensitivities and ANOVA F-scores plotted over time for both sensors. Both classifiers showed equivalent generalization performance of approximately 82% correct single trial predictions.

SVM classifier trained on the balanced dataset achieved a comparable accuracy of 76.07% correct predictions (mean across 100 subsampled datasets), which is a slightly larger drop in accuracy when compared to the 80.8% achieved in the original study [refer to table 3 in 54].

Importantly, these results show that an analysis using statistical classifiers can provide stable results that depend not on a particular implementation, but rather on the methods themselves. Moreover, the example pressented here demonstrates that the integrated framework of PyMVPA can achieve these results with much less effort than was needed in the original study.

3.4 FMRI

In the past decade fMRI became the most popular brain imaging modality among researchers interested in the neural correlates associated with cognitive processing. Three reasons why fMRI is favored over other modalities include the fact that fMRI provides high spatial resolution, it is non-invasive, and data acquisition is relatively easy. A variety of experimental designs and analysis approaches have been created to answer specific neuroscientific questions using fMRI. The following section provides a brief overview of the standard approach to analyzing fMRI data and covers underlying assumptions, promises, and shortcomings. Also included in this next section is an analysis aimed at discovering category specific processing of visual data in the brain. The standard approach is compared with decoding methods proposed in this dissertation.

3.4.1 Introduction

Analysis Methodologies The de facto standard analysis method used to analyze fMRI data (and to some extent E/MEG) is statistical parametric mapping (SPM) [12]. This approach defines two aspects of neural signal processing: *identification* and *integration*. The purpose of identification is to "identify functionally *specialized* brain responses" [12], *i.e.*, to detect regions of the brain (at the level of neural networks) where a statistically significant portion of the neural signal can be well described by known experimental factors. The emphasis on identification segregated from integration significantly simplifies the analysis of fMRI datasets by considering each spatial location (*voxel* in MRI terms) to be independent of any other. This feature makes SPM a type of *mass-univariate* analysis.

The goal of SPM is to make statistical inferences about spatially extended regions. SPM relies on the conjoint use of the general linear model (GLM) and Gaussian random field (GRF) theory (as well as other methods such as Bonferonni correction). The GLM

part of the analysis is used to estimate the parameters of the model per each spatial location in order to explain observed data. The GLM tests the null-hypothesis that there is no effect of the experiment on that location. It is important to emphasize that the GLM relies on the definition of the forward model of the signal, and is therefore limited to linear modeling of a signal. However, the value of using forward modeling of the BOLD fMRI signal is still an open question (as it will be highlighted in Section 4.2). Despite the requirement of forward model specification, being a univariate method, the GLM is relatively immune to misspecification of the forward BOLD model for a wide range of experimental designs. Some *tricks* (*e.g.*, addition of first temporal derivative of the model variables) will improve the identification performance of GLM by accounting for sources of variance that are not explicitly modeled.

Unfortunately, being a type of mass-univariate analysis, the GLM does not account for any covariance or causal structure of the signal (besides local region covariance in some SPM implementations). Therefore, actual identification is limited to the regions with uniform activation across experimental conditions, and cannot distinguish the patterns of activity at the neural network level. Such disregard of the covariance structure is often acceptable with judicious experimental design where experimental factors are carefully crafted and well controlled. Precise control of the neural responses, though, can be very difficult in the cognitive tasks, since the effect of many top-down processes (*e.g.*, attention, memory) are difficult to control. However, top-down processes being a pervasive influence in most cognitive tasks, can not be neglected.

A side-effect of the mass-univariate approach is the need to use a statistical correction method to resolve multiple comparison problems that ensue when making inferences over a volume of the brain that involves tens or hundreds of thousands of voxels. Because often the goal of a GLM analysis is to detect relatively extended brain regions, low-pass spatial filtering is a common preprocessing step to boost SNR and

hence to decrease the number of false positives in the identification task. Indeed it is a valid preprocessing step under assumptions that neuronal population within regions of interest are uniformly engaged within the experimental conditions of interest. But some brain areas, (*e.g.*, high-level visual object processing regions such as fusiform gyrus) are known to be responsive to a variety of stimuli under the right conditions. For such cases, mass-univariate methods lack power to account for or discriminate between spatial patterns of activity. Furthermore, univariate methods are incapable of exploring any *similarity structure* among different stimuli belonging to the same category.

Standard approaches based on the GLM for identification of potentially selective brain regions are neither cross-validated nor optimized for performance [57, 58, 59]. In a typical block design study, the GLM is used to fit selective time series, often accounting for less than 50% of the total variance, despite statistically significant excursions from a predefined baseline (note significance of a statistical test does also imply a high degree of variance accounted for in the *fit* of the linear model). This fact allows for considerable indeterminacy regarding the *identification* of the condition, *selective* voxels, or regions. A common conceptual error that occurs when using this type of method is believing that the strength of the regression coefficients reflect *selectivity* when they are merely indicating the presence of brain tissue activation conditional on the presence of the stimulus exemplar p(VOXEL|CONDITION$>$ *chance*) what many in this field have called *selective* or even *specific* (*i.e.*, uniquely selective). Furthermore, because a bio-physiological model of the BOLD fMRI signal does not exist, a stronger BOLD response does not necessarily relfect *selective* processing at a given location.

Despite these limitations, SPM has been used to support various scientific hypotheses about various *specialized* functional modules within the brain with respect to a given task. However, as noted earlier, reliance on the identification of active brain regions means that the integrative functioning of brain regions can not easily be explored or explained. The authors of SPM, suggest that any aspects of *integration* would need to

be tackled with "a different set of *multivariate* approaches that examine the relationship among changes in activity in one brain area others". Multivariate methods were not considered for the following reasons: "(i) they do not support inferences about regionally specific effects, (ii) they require more observations than the dimension of the response variable (i.e. number of voxels) and (iii), even in the context of dimension reduction, they are less sensitive to focal effects than mass-univariate approaches"[12]. Instead of approaching multivariate methods, which would be partially or fully capable of addressing the aforementioned concerns, so called *effective connectivity* modeling was proposed to address *integration* aspect. Effective connectivity "refers explicitly to the influence that one neural system exerts over another" [12]. To perform this type of analysis, some regions of interest (ROIs) are selected and then the interactions among them are determined. Conventionally, *specialized* brain regions, which are first detected with SPM, are taken to be the nodes in a constructed graph, the edges of which are either provided based on prior assumptions, or discovered via exhaustive or heuristic search. Usually such graph modeling also requires a preliminary *representative time series* extraction procedure to provide a single time series for each node in the graph. Such a step further sacrifices the information at the level of neural networks. So, by design, the explicit separation of *identification* and *integration* during the processing of neural data makes it unlikely that interactions at the level of neural networks will be detected, and the most that can be expected is the discovery of coarse level interactions for "specialized" systems. Consequently, whatever information about *similarity structure* that may exist will not be available.

To summarize, the GLM-enforced separation of *identification* and *integration*, while appropriate for the analysis of data from experiments with large, consistent (across trials), and uniform gross-effects, is not applicable to all experimental designs. For this reason, other methods of analysis, particularly multivariate techniques, may provide a

good alternative. Offered here is a *decoding* approach which relies on constructing a reliable mapping between fMRI neural data and experimental variables.

Category Specific Processing There is considerable debate in the neuroscience community concerning how information about visual objects are processed in the brain. Some researchers [57, 59, 60] have proposed that the representation of some visual objects are localized in specific brain tissue (*e.g.*, fusiform face area (FFA) for "faces" processing [60]). However, other work in this field [61] indicates that FFA may code more general stimulus properties (*e.g.*, "expertise").

Other perspectives reject the notion of localized function Haxby et al. [25] and suggest instead that object identity is coded in a distributed way across the brain, somewhat like a relief or topographic map, so that a particular pattern of response is associated with a specific object type or token [25]. More recent work along these lines [26] indicates that the category specific codes may actually be combinatoric, allowing the efficient reuse of voxel information (*e.g.*, neural populations) across different types or tokens.

Claims that areas of the brain are specific to a given stimulus category are often derived by the comparison of levels of activation between two or more conditions. So, in [60], the difference in detectability between faces in a pre-localized FFA is at its maximum around 2% above baseline, while for other objects (except whole humans, headless bodies, or animal heads) the difference is closer to 1%; a factor of 2 in detectability which, indeed, is quite impressive. Proponents of this view maintain that although other areas of the brain might detect the presence of a face stimulus, these other areas are not uniquely face selective as is the FFA. However, supporters of this claim are not distinguishing face detection from face identification. Identification is the ability to label a given stimulus as belonging to a specific category. The degree to which errors occur can be captured in a classic confusion matrix which delineates both the discriminability and similarity of

various object tokens [62]. Detectability is related to identification to the extent that it is necessary for identification, but it is not sufficient to make specific category assignment.

Luce [63] has formalized this relation in what is known as the *Luce Choice Axiom*[7]:

$$P(i|v) = \frac{w_{(i|v)}}{\sum_j w_{(j|v)}} \qquad i = 1, \ldots, n, \tag{3.1}$$

where choice identification $P(i|v)$ depends on conditional probabilities $w_{(i|v)}$ of the category i given the voxel or feature v, relative to all other potential category choices. This type of calculation can be made with strength or detectability measures such as regression coefficients. When it is done correctly (*e.g.*, Haxby et al. [25]), voxels involved in face identification were found to be distributed over much of inferior temporal lobe.

As described earlier in this thesis, work in visual object recognition that relies on fMRI measures often uses the activity of voxels above some baseline activity to infer selectivity. Although this may seem like a benign choice, perseverates a view that selectivity of cortical tissue can be measured unambiguously with normalized BOLD responses. The problem with these measures are actually two-fold. First, when using the GLM, a detection is calculated that relies on a contrast between a particular stimulus condition and one or more others. This means that voxel activity is used basically as a regressor. As such, the GLM is not diagnostic or able to identify category membership that is conditionally dependent on that feature. Worse, the intensity of voxel activation, which covaries with the strength of the coefficient, is confounded with the voxel's location. Moreover, if blood flow to cortical regions actually signaled the amount of "work" or "energetics" involved in processing states, then BOLD intensities that were not at peak levels might indicate more efficient processing than those at peak levels. This possibility means that selectivity measures that are designed towards diagnostic

[7] http://en.wikipedia.org/wiki/Luce's_choice_axiom

identification, rather then towards strength of association, could downplay the importance of voxels that had lower intensity but which were actually more reliable predictors. This poses a problem for cognitive neuroscientists who are interested in finding the brain mechanisms that underlie correct classification or identification of a stimulus and not simply in detecting where the greatest neural response occurs.

The behavioral response of a subject who is asked to identify a masked stimulus [64] will produce a standard psychophysical function that reflects changes in the ability of the subject to correctly identify the stimulus as the mask is changed. However, for the associated fMRI data, a GLM analysis of voxel values can identify only those voxels that are associated or most similar to systematic variations of the independent variable (*e.g.*, the mask). Thus, the GLM does not provide the type of voxel identification function most neuroscientists are seeking. This is where statistical learning theory, and statistical classifiers in particular, can be of considerable benefit.

Classifier-based Analysis There has been enormous progress over the last few years in the development of statistical classifiers. Recent work often focuses on problems having extremely large feature spaces ($>$ 100k to 1M) and poor signal/noise ratio, often with remarkable out-of-sample generalization [65]. That why these methods may be particularly useful in analyzing fMRI data. The BOLD signal is known to have notoriously low signal/noise gain and to reflect a mixture of various neural signals and field potentials. There is also growing evidence that the underlying distribution is non-Gaussian [66, 67, 68], which means that parametric classifiers may underestimate valid signal excursions and make approaches *e.g.*, Haxby et al. [25] biased toward conservative generalization estimates. In fact, when Hanson et al. [26] compared Haxby's nearest neighbor classifier to neural network classifiers, a conservative cross-validation estimate showed improvements in generalization by as much as 30-40%. Much of the improvement obtained with statistical classification can be attributed to a

more appropriate classification function (*e.g.*, specific nonlinearity) as well as feature weighting and/or selection [69].

In order to discover features (areas of the brain) that are responsive to specific stimulus types it is necessary to focus on whole brain classification. To date of [28] researchers in the visual object recognition domain have used SVM (see [70]) in more generic "brain reading" contexts and, with the exception of [31] who used single scans to filter support vectors, no one has attempted whole brain, single scan classification, focusing instead on regions of interest such as the inferior temporal (IT) lobe. On the other hand, a number of researchers in the object recognition field have found evidence that "object-selective" or "face-selective" areas of the brain exist outside of IT [71, 72, 73]. However, it still remains unclear whether these "selective" areas uniquely identify certain object types. Do these other brain areas, that are neither FFA or parahippocampal place area (PPA), carry *discriminative* information about exemplars like faces or houses? At level of category structure do discriminative areas function?

As the first step toward answering these questions, Section 3.4.1 describes how statistical learning methods can be used to identify areas engaged in object processing, and to associate a level of semantic processing to those areas (*e.g.*, like that used to distinguish between animate and inanimate categories). Moreover, the existing claim that some brain areas, such as fusiform face area, are category *specific* is challenged In Section 3.4.3 statistical learning methods are used to demonstrate that category-*specific* areas do carry discriminative information about nonpreferred stimuli.

3.4.2 Identification: Multiple Categories Analysis

Data from a single subject who participated in a study published by Haxby et al. [25], which has been reanalyzed by a number of researchers since the original publication [13, 26, 28], serves as the example fMRI dataset for the unimodal analysis of fMRI presented

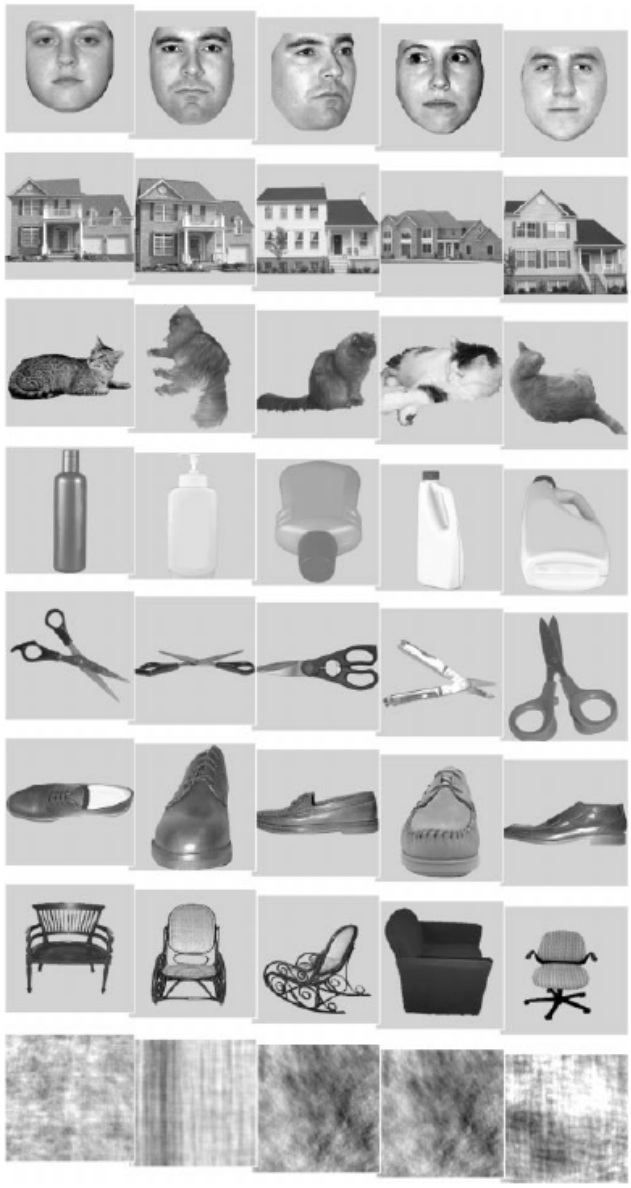


Figure 3.5 Stimuli exemplars from [25].

in this dissertation. The dataset itself consists of 12 runs. In each run, the participant passively viewed greyscale images (see Figure 3.5) of eight object categories, grouped in 24 s blocks, and separated by rest periods. Each image was shown for 500 ms and was followed by a 1500 ms inter-stimulus interval. Whole brain fMRI data were recorded with a volume repetition time of 2500 ms, so that brain activity associated with a given stimulus block was captured by roughly 9 volumes. For a complete description of the experimental design and fMRI acquisition parameters see Haxby et al. [25].

First, raw fMRI data were motion corrected using FLIRT from FSL [74]. All subsequent data processing was done with PyMVPA (see Section 5). After motion correction, linear detrending was performed for each run individually. No additional spatial or temporal filtering was applied. For the sake of simplicity, the dataset was reduced to a four-class problem (*faces*, *houses*, *cats* and *shoes*). All volumes recorded during any of these blocks were extracted and voxel-wise z-scored. This normalization was performed individually for each run to prevent any information transfer across runs.

After preprocessing the same sensitivity analysis used for all other data modalities (presented in previous sections of Chapter 3) was applied to this dataset. Here, only a SMLR classifier was used (6-fold cross-validation, with two of the twelve experimental runs grouped into one chunk, and trained on single fMRI volumes that covered the whole brain). For comparison, a univariate ANOVA was computed again for the same cross-validation dataset splits.

The SMLR classifier performed very well on the independent test datasets, correctly predicting the category for 94.7% of all single volume samples in the test datasets. To examine what information was used by the classifier to reach this performance level, region of interest (ROI) based sensitivity scores were computed for 48 non-overlapping structures defined by the probabilistic Harvard-Oxford cortical atlas [75], as shipped with FSL [76]. To create the ROIs, the probability maps of all structures were thresholded at 25% and ambiguous voxels were assigned to the structure with the

higher probability. The resulting map was projected into the space of the functional dataset using an affine transformation and nearest neighbor interpolation.

In order to determine the contribution of each ROI, the sensitivity vector was first normalized (across all ROIs), so that all absolute sensitivities summed up to 1 (L_1-normed). Afterwards, ROI-wise scores were computed by taking the sum of all sensitivities in a particular ROI. The upper part of figure 3.6 shows these scores for the 20 highest-scoring and the three lowest-scoring ROIs.

The lower part of the figure shows dendrograms from a hierarchical cluster analysis on relevant voxels from a block-averaged variant of the dataset (but otherwise identical to the classifier training data). For SMLR, only voxels with a non-zero sensitivity were considered in each particular ROI. For ANOVA, only the voxels with the highest F-scores (limited to the same number as for the SMLR case) were considered. For visualization purposes the dendrograms show the distances and clusters computed from the average samples of each condition in each dataset chunk (i.e, two experimental blocks), yielding 6 samples per condition.

The four chosen ROIs clearly show four different cluster patterns. The 92 selected voxels in temporal occipital fusiform cortex (TOFC) show a clear clustering of the experimental categories, with relatively large sample distances between categories. The sensitivity plot (upper part of Figure 3.6) shows that univariate statistics undervalued the voxels within TOFC and LOC inf, although those are the areas which carry category-specific information. The pattern of the 36 voxels in angular gyrus reveals an animate/inanimate clustering, although with much smaller distances. It is an interesting result given that clustering was done in an unsupervised fashion. However, it is a good demonstration of how a multivariate technique can reveal the presence of information at this level of semantic specificity.

The largest group of 148 voxels in the frontal pole ROI seems to have no obvious structure in their samples. Despite that, both sensitivity measures assign substantial

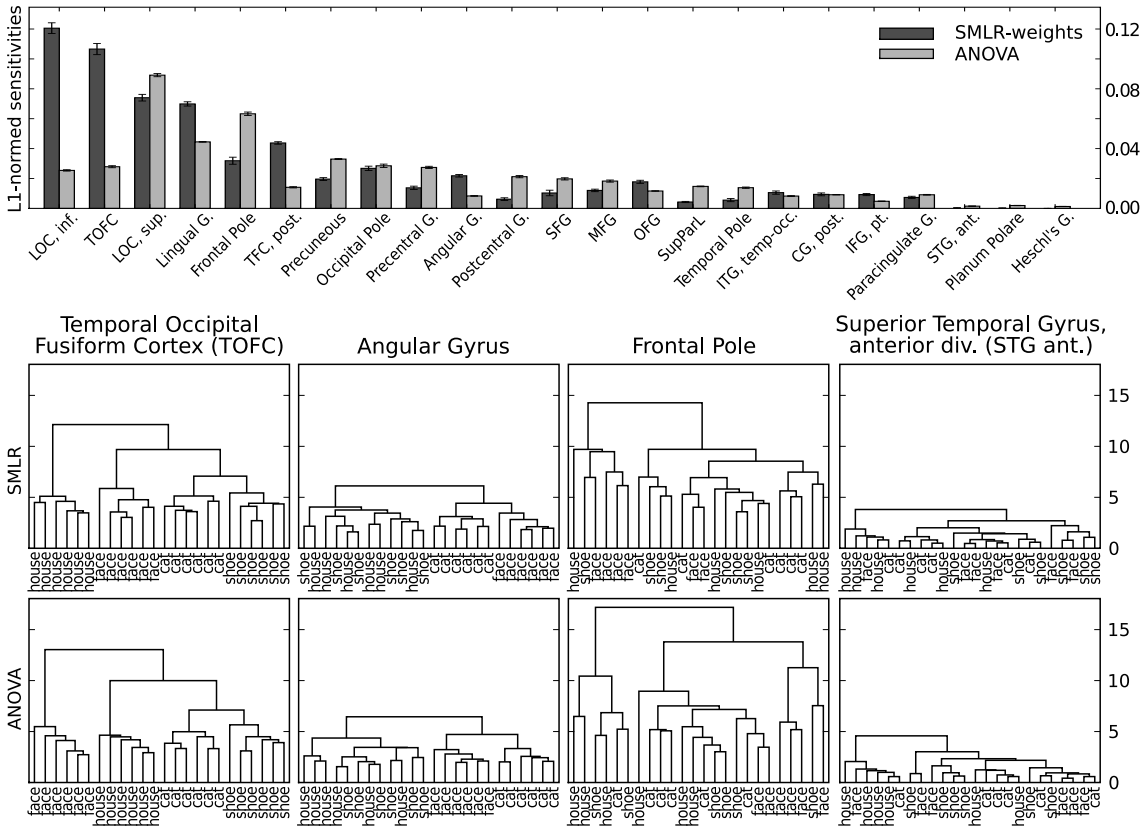


Figure 3.6 Sensitivity analysis of the four-category fMRI dataset. The upper part shows the ROI-wise scores computed from SMLR classifier weights and ANOVA F-scores (limited to the 20 highest and the three lowest scoring ROIs). The lower part shows dendrograms with clusters of average category samples (computed using squared Euclidean distances) for voxels with non-zero SMLR-weights and a matching number of voxels with the highest F-scores in each ROI.

importance to this region. This might be due to the large inter-sample distances visualized in the corresponding dendrogram in figure 3.6. Each leaf node (in this case an average volume of two stimulation blocks) is approximately as distinct from any other leaf node, in terms of the employed distance measure, as the semantic clusters identified in the TOFC ROI. Finally, the ROI covering the anterior division of the superior temporal gyrus shows no clustering at all, and, consequently, it is among the lowest scoring ROIs of both measures. On the whole, the cluster patterns from voxels selected by SMLR weights and F-scores are very similar in terms of inter-cluster distances.

Given that these results rely on data obtained from a single participant, no far-reaching implications can be drawn. However, the distinct cluster patterns provide an indication that different levels of information encoding could be addressed in future studies. Although voxels selected in both angular gyrus and the frontal pole ROIs do not provide a discriminative signal for all four stimulus categories, they nevertheless provide some disambiguating information which the classifier was able to detect. In angular gyrus, this seems to be an animate/inanimate contrast that can also differentiate between animate stimuli belonging to two categories. Finally, in the frontal pole ROI the pattern remains unclear, but the relatively large inter-sample distances might indicate a differential code of some form that is not closely related to the investigated level of semantic stimulus categorization.

3.4.3 Specificity Assessment: Face vs. House Debate

To address the question of whether or not a face-*specific* area exists, this section presents an analysis of a decoder capable of discriminating between only face and house categories. Although a number of classifiers were tested, the analysis presented here focuses on Support Vector Machines (SVMs) [77]. As noted earlier, SVMs have many desirable properties: they can learn even in huge ($>$1M) feature spaces, they produce a

unique solution by constraining problem formulation, they generalize well with feature spaces many orders larger than the data sample size, they can be highly robust over significant levels of noise, and finally, they can learn subtle distinctions near the separating hyperplane which can disambiguate each category sample. SVM does this all without the cost of learning the complete distribution of each category (although this will turn out to be a disadvantage for visualization).

To address the question of brain area identification, the whole brain was used as input ($\approx$ 40K voxels – white matter was used as a noise background to increase the reliability of estimates). The data was then incrementally searched for voxels that uniquely identified either FACE or HOUSE stimuli. This was done exhaustively per subject per scan and cross-validated on independent sets of data. In order to increase the generalization of this approach, data from two separate object identification experiments were used. In both studies, subjects performed judgments on the two key object types (FACE, HOUSE) in two different cognitive tasks. The first task was the same as in the previous section (*i.e.*, [25]) with a focus on the stimuli necessary for the identification of FACE and HOUSE, so that only blocks of collected for those two conditions were considered in the analysis. In the second task, memory demand is reduced by using a simple perceptual judgment in which high contrast black and white stimuli were presented in an oddball task (task 2 is designated as the OB task). In the OB task, trials consisted of a simultaneous presentation of 3 stimuli (either 3 FACEs in FACE trials or 3 HOUSEs in HOUSE trials) in different orientations with the subject being asked to identify the member of the triad that was different. In both tasks subjects achieved behavioral accuracy rates identifying objects above 80% correct. In both cases, full brain (approximately 40000-50000 voxels) data was collected, with 144 samples (77 of each type) in task 1 and 200 samples (100 of each type) in task 2. Both experiments used a block design and data provided to the analysis consisted of 7 trials in task 1 and 17 trials in task 2 (3 of the original 20 trials were eliminated to reduce autocorrelation effects). Leave-two (one

of each category) blocks-out strategy was used to assess the generalization performance. Classification was done with single scans or equivalently single TRs – c.f. [26]. In order to achieve the lowest possible error in generalization, recursive feature elimination (RFE) was performed (see [65]). This approach has the advantage of detecting the specific object identification brain areas by harvesting the most sensitive voxels after training and doing subsequent retraining on this sensitive, but smaller set. This process was continued until there are no more voxels left to test and therefore, was exhaustive over the single scan brain voxel set.

To avoid potential cross block contamination of hemodynamic BOLD signal and therefore accidentally creating a generalization bias, scans at the beginning and the end of all blocks (category or rest conditions) were discarded. Whole blocks were also routinely held out, and tested against single TRs from those blocks (holding out all other TRs in that block until sampled later). This approach reduced any upward bias from possible correlation within blocks. A 20 second rest block was also used in both experiments to further prevent any temporal contamination or confusion between category responses. In order for the SVM classifier to operate on the data, raw voxel data had to be converted so that values fell within [-1,+1] range. Two possible conversion schemes were tested: scaled percentage change relative to baseline, and z-scores relative to the baseline (rest condition blocks). SVMs trained on z-scores provided better generalization on chosen test cases and so these were chosen for further analysis. Each 3D scan (brain voxels only) containing roughly up to 40,000 voxels was used as an input sample for the classification, in addition to a label of the stimulus condition (FACE or HOUSE) presented during that scan. Subject data was submitted to a soft-margin SVM with an average per step backward feature elimination of 1000-500 voxels (approximately 12% to 15% exponential removal rate until 100 voxels was reached and then removal of one voxel at a time until only one voxel was left) . At each step, a voxel's sensitivity within trained SVM was used as a basis

Table 3.1 Minimal Error and Associated Voxels Count.

Gen/Subj	NB1	NB2	NB3	NB4	NB5	OB1	OB2	OB3	OB4	OB5
VOXELS	200	104	117	120	34	378	125	81	313	225
ERROR	0.0	6.8	0.0	0.0	6.9	0.0	2.7	6.7	3.0	0.0

for keeping or eliminating that voxel. Voxels with the smallest weights on each step were eliminated.

In order to obtain an unbiased estimate of SVM generalization performance each dataset was split into training and testing datasets. Similar to [26], a N-1 block bootstrap procedure was implemented. Specifically, for each training set a single block from each category (FACE or HOUSE) was taken out for testing, leaving B-1 per category used for training. All possible combinations of testing blocks from the two categories were taken, making a total of 144 NB (12x12 blocks) and 100 OB (10x10 blocks) training/testing datasets[8]. Each training/testing dataset proceeded through RFE independently and performance was averaged over all classifiers/bootstraps to obtain a generalization estimate for a given subject/SVM. Shown in Figure 3.7 is the average error of the generalization as a function of the voxels remaining in the training set. Each line shows a single subject's performance on out of sample pairs of HOUSE and FACE exemplars. The color of the line indicates the group of subjects either in NB (red) or OB (green) tasks, note that near the minimum they significantly overlap, with the oddball task starting at a lower error on initial learning[9].

[8]For two subjects, one in the OB task and the other in NB, a single block in each case was found to be corrupted, and for those cases the number of bootstrap opportunities were 81 and 121 respectively.

[9]One possible reason for the advantage in the OB task could simply be the difference in scanner strength which was 1.5T for the NB task and 3T for OB task. Another possible explanation is that the OB task actually required a category judgment, while the NB task required only a simple stimulus identification. This would change the behavioral contrast between stimuli. In either case, a parametric experiment that varied the difficulty of the categorization judgment might resolve this intercept difference.

In table 3.1 is shown the exact minimum error and associated voxels remaining at that point. The minimum average error for all subjects in both tasks indicated nearly 98% correct on out of sample cases with 3 subjects in the NB task and 2 from the OB task showing ZERO generalization error.

Brain Areas Identification Which areas of the brain are uniquely required for the identification of FACEs and HOUSEs? In a fashion similar to that described in previous sections, constructed classifiers were examined by doing a sensitivity analysis that effectively asked what a voxel's contribution to the classification error was. If category error significantly changes when the voxel is removed, then it can be inferred that this voxel is contributing in a proportionally diagnostic way to the identification of the FACE or HOUSE category.

All brain volumes (tasks, subjects) were first registered using the following steps: (1) a subject's sample bold scan, anatomical, and MNI anatomical were skull stripped using BET from FSL tools [76], (2) the stripped anatomical was co-registered to the stripped MNI using FLIRT (from FSL) to obtain an anatomical to MNI transformation, (3) the stripped anatomical was co-registered to the stripped BOLD using FLIRT to obtain an anatomical to BOLD transformation, and finally (4) the anatomical was transformed into BOLD space using anatomical to BOLD transformation for easy visualization of activation patterns.

In order to localize the FFA and PPA for analysis, the standard method in the field for localizing category responsive voxels was used. This involved first finding voxels that significantly responded to FACE $>$ "other non-FACE objects" or HOUSE$>$ "other non-HOUSE objects". In the case of the NB task, the original FACE and HOUSE masks in Haxby et al's [25] study were used. Those masks ranged from 20 voxels to about 100 voxels. In the case of the oddball study, independent localizer scans were used in standard GLM contrasts for FACE$>$HOUSE and HOUSE$>$FACE. Although seemingly

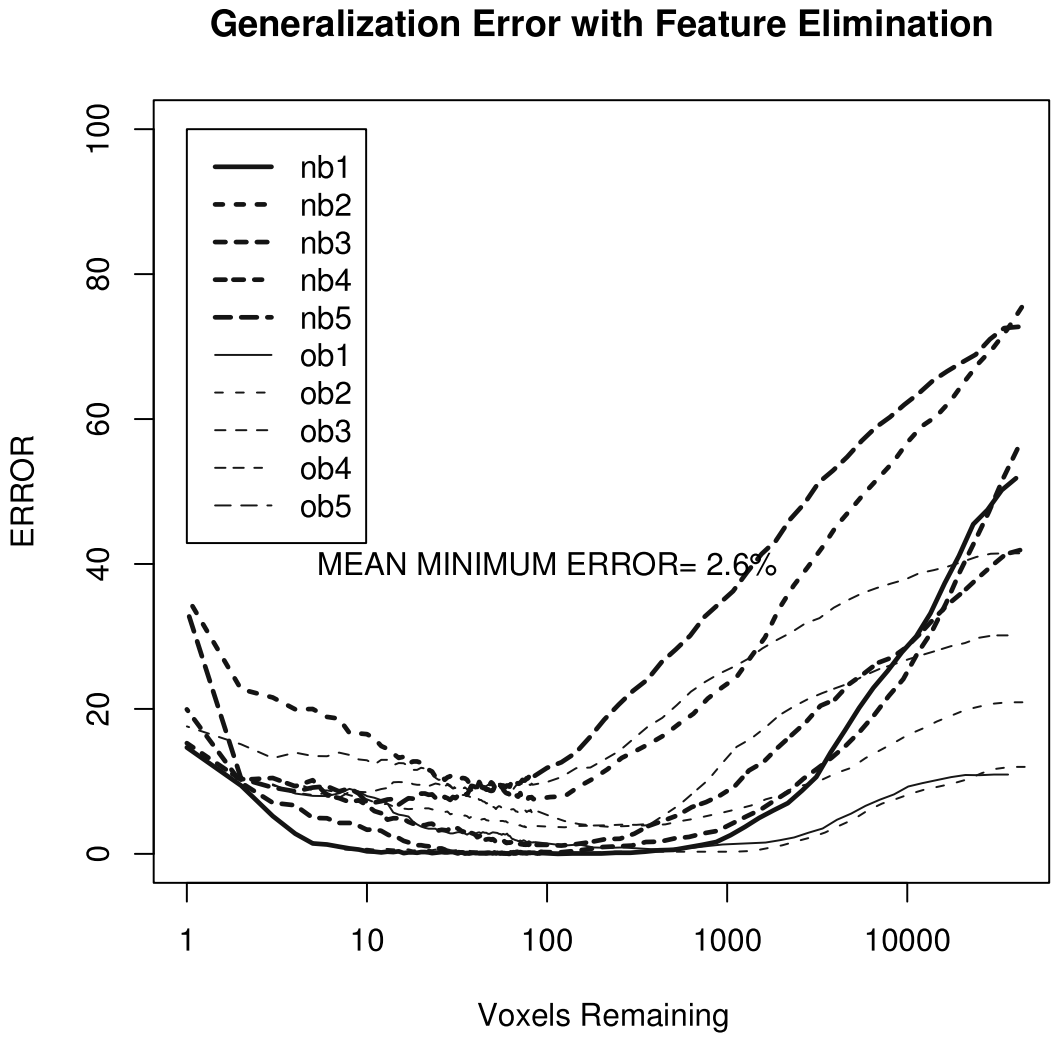


Figure 3.7 Generalization error for all ten subjects on single scans for both kinds of stimuli held out from a training exemplars. Recursive voxel elimination produced a minimum for half the subjects at zero error, while over all subjects on single scans, across both tasks there is nearly a 98% correct generalization on unseen scans.

tautologous, this is nonetheless the standard procedure within the literature to establish selective ROIs for subsequent testing. Recently, there has been considerable debate about this method [78, 79], but nonetheless, this procedure remains the primary way of determining *candidate* voxels for more diagnostic approaches. In the followup analysis FFA and PPA masks were intersected with the calculated sensitivity masks to determine the amount of voxel overlap. The difficulty in identifying the FFA in particular, is that there can be considerable variability in the size of masks across subjects. In the original paper [57] that argued for the existence FFA, only 12 of the 16 subjects actually had FFA activity. Despite its limitatons, identification of the FFA by GLM contrasts of localizer scans reflects the state of the art in the neuroimaging field for selecting candidate voxels.

Sensitivity Analysis There are a number of possibilities for identifying diagnostic brain areas. An obvious choice might be to use all of the support vectors themselves. However, using vectors that are strictly in the margin may be inappropriate in characterizing the diagnostic brain areas for HOUSE and FACE inasmuch as they might represent brain volumes that are near the separating surface, and therefore unrepresentative of the brain response to either HOUSE or FACE stimuli. The other possibility is the NON-support vectors (NSV) described by LaConte et al. [31]. These are vectors that are distributed beyond the support vectors and although some may be prototypical of the brain response, unfortunately many will not be. In fact, in general even assuming a Gaussian spread of the NSVs only a small minority will be typical or "best" members of the category. Because SVM optimizes a margin between the two categories, the actual distribution of members of the categories is effectively ignored. Hence, ironically, the same property that makes SVM an excellent candidate for classification in high dimensions is the one that also makes it tricky to visually interpret. For this reason, a visualization approach between these two extreme possibilities was chosen. Effectively it implements a sensitivity/perturbation approach (*e.g.*, [26]), which

measures the error for a given category (FACE, HOUSE) when the voxel is present verses when it is removed. Specifically, to estimate SVM-based sensitivity, we used one of the simplest criterion that have been proposed [41, 65, 80], which is simply the reciprocal of the separating margin width $W = \|w\|^2$, where $w = \sum_i \alpha_i y_i x_i$. Minimization of this criterion leads to maximization of the margin width. In the case of Linear SVM, the squared values of the separating plane normal coefficients (i.e. w_i^2), as stated, effectively correspond to the change of the criteria *W* as if the voxel *i* is removed. Therefore, the classifier is less sensitive to the features with low w_i^2. During recursive feature elimination we sequentially eliminated features with the smallest w_i^2. Additionally, in order to increase diagnostic selectivity we derived weights for the FACE category by using only FACE SVs and for the HOUSE category by using only HOUSE category SVs. Thus, higher voxel values tended toward typical regions in the classification space for the SV appropriate category. We will refer to these direction selective voxel coefficients as *diagnosticity* of the voxel direction for or against HOUSE (in blue) and FACE (in red). Shown in Figure 3.8 are diagnosticity measures for two typical subjects (one from NB and one from OB). A non-parametric method for thresholding these diagnosticity distributions at $p<0.01$ was used for each subject, taking into account the non-Gaussian aspect of the underlying diagnosticity distributions (c.f. with Type 4 method constructing empirical cumulative distributions; linear interpolation of the cdf [81]). Slices are shown for each subject with specific voxel clusters. Note that SVM finds fairly contiguous regions, despite the specific bias not to. In fact, at lower thresholds (0.05) the diagnosticities begin to fractionate through the whole brain. Unlike a regression analysis (*e.g.*, GLM) which finds "selective" or detection areas, these voxel patterns are unique identification areas in that they are contributing to a correct classification of one category against the other, and moreover, they have been cross-validated in independent data sets so that the likelihood is high that they will generalize to other unseen cases of either FACEs or HOUSEs.

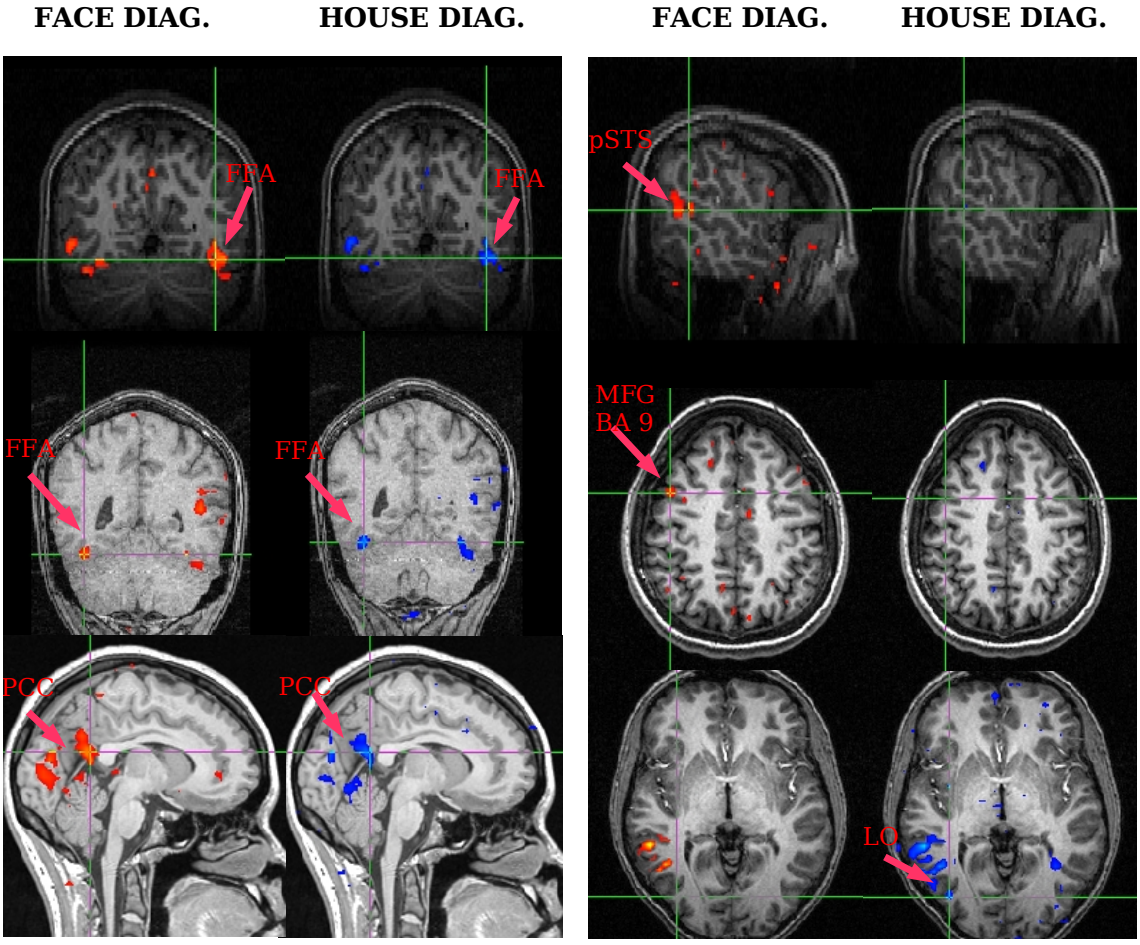


Figure 3.8 Voxel sensitivities for brain areas that are diagnostic for FACE (red and on left of each panel) and for HOUSE (blue and right of each panel). In the top left corner we show FFA (fusiform face area as measured by a localizer task) in the first paired panel (same subject, same slice, same task) which shows subject NB1. Below that is subject OB3 also showing FFA in both slices. On the left bottom panel we show PCC which is diagnostic for HOUSE and FACE which both show sensitivity (NB2). The next three paired images show distinctive areas between diagnostic areas of FACE and HOUSE. The next panel above on the right shows pSTS (posterior superior temporal sulcus) which is diagnostic for FACE against HOUSE (subject OB1), the next panel below shows another FACE diagnostic areas (MFG BA9, subject OB5), and finally in the last paired panel on the right we show a HOUSE diagnostic area (LO) for subject NB5.

The questions posed at the beginning of this section may now be addressed by assaying the above threshold identification areas, in order to determine whether (1) there are unique identification areas for FACE and HOUSE, and (2) whether the FFA or PPA are carrying other discriminative information about other nonpreferred stimuli (in this case either HOUSE or FACE respectively).

Brain Areas Identifying FACEs and HOUSEs Figure 3.9 shows all areas harvested that were common (intersection set[10]) to all subjects in both tasks that were either highly diagnostic of FACEs (2a) or of HOUSEs (2b) based on the sensitivity/diagnosticity analysis as described before. The barplot shows the percent of voxels associated with each area at the 0.01 threshold using the nonparametric methods. The total in each bar can be computed by multiplying the percentage against the total number of sensitivities above threshold in each category (FACE=363 and HOUSE=358), so for example about half the voxels in both HOUSE and FACE, about 150 each fall into the FFA. Note that these areas are based on the best cross-validated single scan classifiers that were nearly 98% correct on out-of-sample exemplars. Hence the areas identified under these constraints are not based on the usual "object selective" interpretation and therefore not subject to the resultant ambiguity with methods that are based on similarity or association. They are in fact, albeit in a probabilistic sense, necessary and sufficient for identification of FACEs or HOUSEs. It is important at this point to clarify that the present classifiers could have found single areas or single voxels as predictive of a single category. In fact, linear methods tend to be biased towards using single dimensions (voxels in this case) to minimize classifier error especially if area correlations tend to be

[10] A very conservative harvesting was used to include only voxels that were in all ten subjects voxels that were above $p<0.01$ therefore appeared in both tasks. Also, sensitivities were initially averaged over all generalization runs in order to increase the sample power of the voxels sets per subject. In any case, there was significant overlap (64%) of the same voxels across generalization runs and this tended to covary with the minimum generalization error reached for that classifier. It is appropriate to average across the runs, since any differences in voxel sets are due to sampling error in classifier estimation and data noise.

small between features or voxels. If there are large correlations between voxels these could be minimized by using a Neural Network which can decorrelate features as it classifies. Nearest Neighbor classifiers such as used by Haxby et al [25], in fact are more biased towards sets of features since their similarity increases incrementally with more common overlap of voxels. Consequently, it is possible to begin answering the question posed in this section: Are there other areas of the brain that are diagnostic for the identification of FACEs or HOUSEs? Do other areas of the brain carry discriminative information other than FFA and PPA? Clearly from the present analysis the answer is "yes".

Overall, the diagnostic profile of relevant areas for both FACEs and HOUSEs are somewhat similar, but not identical. Despite the appearance of distinct areas between categories, these areas also seem to be part of a larger network. In fact, Figure 3.8 shows a selection of representative examples from different subjects across both tasks (OB and NB). In the figure each paired set of figures in a row is the same subject showing the diagnosticities for FACE in red (on the left of each pair) and HOUSE in blue (on the right of each pair). In the first two paired sets we show FFA sensitivity in two different subjects across the two different tasks, in both FACE and HOUSE stimulus presentations. For all subjects fusiform gyrus (and FFA masks overlapped with 90% of the FG voxels for all subjects and both tasks) appeared for both FACE and HOUSE indicating diagnostic value for this area that was neither specialized nor unique. Also for all subjects across the tasks the parahippocampal gyrus (and PPA masks overlapped with more than 70%) and the posterior cingulate cortex (PCC) were also diagnostic of FACE and HOUSE. The next three sets show *distinctive* diagnostic areas, including unique STS sensitivity for FACE but not HOUSE. Other common diagnostic areas include, middle occipital gyrus (including area LO), and middle frontal gyrus (BA 9). In general, a network of areas was identified as diagnostic of these two stimulus types, with FACE having a prefrontal area involved, while HOUSE appeared to involve an area known for visual shape/texture processing.

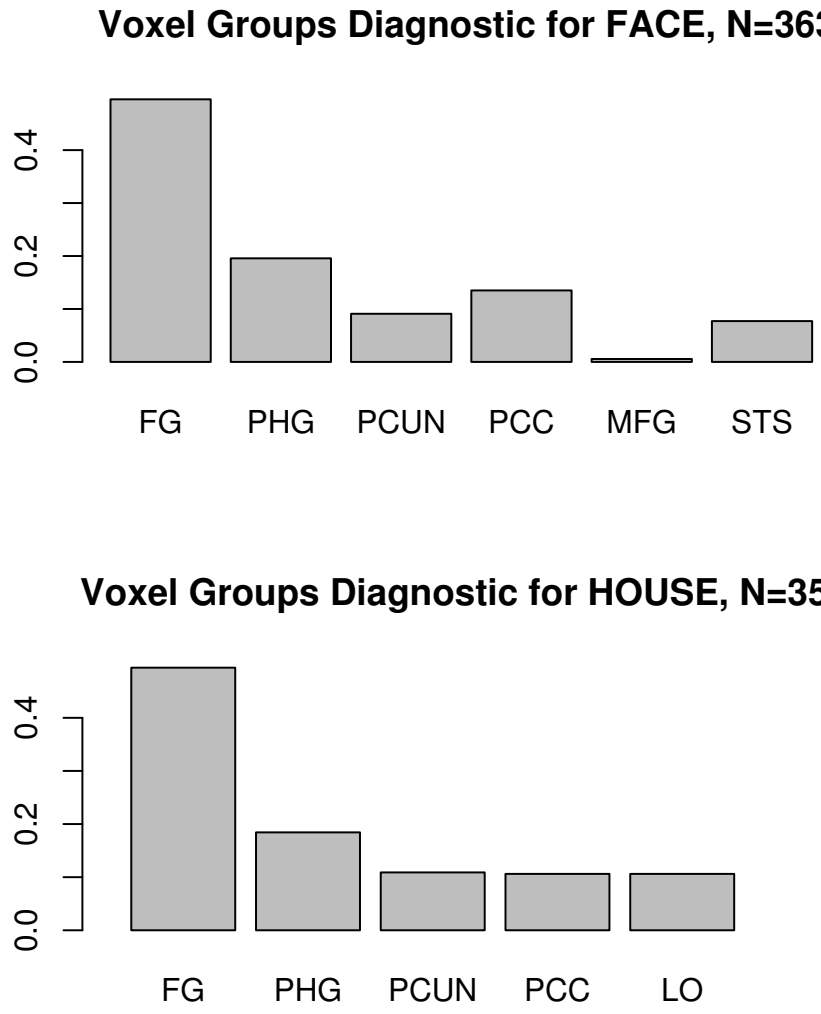


Figure 3.9 Voxel area cluster relative frequencies intersected over all subjects and both tasks. For FACE diagnosticities (top), six areas were identified at $p<0.01$: Fusiform Gyrus (FG), Parahippocampal Gyrus (PHG), Middle Frontal Gyrus (MFG), Superior Temporal Sulcus (STS), and Precuneus (PCUN), with relative frequencies based total number of voxels (N=363) over all areas. For HOUSE diagnosticities (bottom) there were five areas, (at $p<0.01$, N=358) some the same, some different, including FG, Parahippocampal Gyrus (PHG), Posterior Cingulate Cortex (PCC), PCUN, and Lateral Occipital (LO). Note that overlapping FACE and HOUSE areas, include FG, PHG PCUN and PCC. Distinctive areas for FACE included STS, MFG, while for HOUSE distinctive areas included only LO.

Note, that the claim is not that FACE stimuli, do not require specialized functions (*e.g.*, shape/texture) processing. Rather, it seems likely that the areas that have been identified in specialized experiments induce us to use labels that have some lexical familiarity with the presented stimuli and which may be misleading in other contexts where those brain areas are interacting with many other brain areas. This labeling problem and other implications from this study are discussed in the following section.

3.4.4 Discussion

This section of the dissertation attempts to answer a simple question that has been plaguing the object recognition field for the last 10 years: is there a unique area of the brain that is used only to identify faces? Further, is there a unique area of the brain who's sole purpose is to identify houses? Based on the present analysis the answer is "no". To be fair, this statement can be qualified in two ways. First, a claim is not being made about some general property of "combinatorial" coding throughout the brain, but rather the results of the analysis presented here are specific to the inferior temporal lobe and for object recognition processes in particular. Second, a finer resolution of data (see discussion below) might change the results dramatically as more detail within inferior temporal lobe and FFA in particular are made availabe. Nonetheless, the analysis did find a disjoint network of brain areas that are diagnostic for either FACE stimuli or for HOUSE stimuli. This is consistent with Haxby et al's [82] model of face perception which finds the same areas to be involved in a common network for processing faces. The common areas in this analysis include the fusiform face area, specific areas of the lingual gyrus (which showed up weakly in this analysis), the middle occipital gyrus (LO), and precuneus. Distinctive areas for FACE stimuli included pSTS, and middle frontal gyrus (BA9). Distinctive areas for HOUSE stimuli included only area LO. However, for FACEs, there seems to be no evidence in the present analysis that the FFA

is either unique diagnostically nor specialized insofar as it was identified in every subject responding to the HOUSE stimuli.

So this observation would seem to put a rest to the controversy that there are unique areas of the brain that only respond to specific tokens or types. But not exactly. The claim of Kanwisher [60] is complicated by the fact that many areas of the brain seem to be required for identification of these FACE and HOUSE stimuli and the assumption that areas that are *already* selected through an independent localizer (see earlier discussion) tests are uniquely and solely involved in identifying these object types. So to paraphrase Kanwisher, does the FFA and PPA provide discriminative information about other object types other than FACE and HOUSE respectively? The answer based on the present classifiers is "yes". Since most of the voxels identified for either FACE or HOUSE were squarely sitting in the FFA of each subject, this would seem to be definitive evidence that the FFA does involve discriminative information about object types other than FACEs, that is, in this case, HOUSEs. We also must note that the exact overlap of the areas between FFA-FACE and FFA-HOUSE are not exactly the same, despite voxels that are squarely in the same place (see crosshairs), nonetheless this is normative in the field as the FFA will have different locations and shapes across subjects and even within a subject across sessions or experiments will vary in strength, location and shape. Given HOUSEs and FACEs do look different, it is not impossible that the FFA does code these stimuli differently, and a an experiment using fMRI data obtained at a higher resolution might very well produce different distributions of recruited voxels in the FFA for each stimulus, which might be more consistent with Kanwisher's claims. Nonetheless within the state of the art, our localizations of diagnostic cortical areas have no more variability or lack of precision than that which appears in the standard literature. A potentially more difficult issue for Kanwisher is that in our results there were also brain areas that were distinctive for FACE or HOUSE other than FFA or PPA. For example, it would be possible to argue that pSTS is the superior temporal sulcus FACE area, the pSTSFA! We know that this

area is sensitive to “biological motion”, why could there not be a part of it specifically dedicated to the unique identification of FACEs? Other category tests would be very likely to show these areas area not distinctive, but more likely part of a larger network of some kind of FACE identification system. In terms of uniqueness and given the specificity of the claim, that the FFA is claimed to be unique and necessary for FACE identification, then if it were also to be diagnostic for any other category such as HOUSEs, the claim must be refuted. Again, this would seem to lay the matter to rest, there is no FACE area per se, at least in the way that Kanwisher has defined it. Clearly, it is obvious to conclude that there is no area that responds uniquely and solely to FACEs, that could be found in a whole brain assay of a high performance single TR classifier.

CHAPTER 4

MULTIMODAL ANALYSIS OF FMRI AND EEG DATA

There is an increasing number of reported E/MEG/fMRI conjoint studies, which attempt to gain the advantages of a multimodal analysis for experiments involving perceptual and cognitive processes: visual perception [83, 84, 85, 86, 87] and motor activation [83], response anticipation (via contingent negative variation CNV) [88], somatosensory mapping [89, 90], fMRI correlates of EEG rhythms [91, 92, 93, 94, 95], arousal and attention interaction [96], auditory oddball tasks [97, 98, 99], passive frequency oddball [100], illusory figures in visual oddball tasks [101], perceptual closure [102], target detection [103, 104], face perception [105], sleep [106], language tasks [84, 107, 108], self-regulation for EEG-based BCI [109], and epilepsy [110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120].

Researchers who approach multimodal data collection and analysis run into a multitude of problems with regard to the acquisition and analysis of neural data. Examples of these problems will be provided in the first section of this chapter as well as a discussion of the advantages that existing multimodal analysis techniques offer over unimodal analysis approaches. It will be shown how the exploration of new techniques for analyzing components of E/MEG signals, in addition to the use of those already in existence, can be used successfully in conjoint analysis. Thus, this dissertation outlines the problems with existing methodologies and offers a novel multimodal data analysis approach as a solution.

4.1 Multimodal Experiment Practices

When you build bridges you can keep crossing them

– Rick Pitino

Although a strong static magnetic field itself seems to have no effect on the brain electric activity (see [121, chap. 4.2] for review), obtaining non-corrupted simultaneous recordings of EEG and fMRI is a difficult task due to interference between the strong MR field and the EEG acquisition system. Because of this limitation, a concurrent EEG/fMRI experiment requires specialized design and preprocessing techniques to prepare the data for the analysis. The instrumental approaches described in this section are specific to collecting concurrent EEG and fMRI data. For obvious reasons MEG and fMRI data must be acquired separately in two sessions. However, even when MR and MEG are used sequentially, there is a possibility of contamination from the magnetization of metallic implants which can potentially disturb MEG acquisition if it is performed shortly after the MR experiment.

A series of validation studies [122] has been carried out to prove the viability of collecting EEG concurrently with fMRI. These studies use existing design schemes and post-processing to isolate the contaminated EEG signal. Thus the study [122] showed high concordance in signal characteristics between various neuronal activations (steady state visual evoked potentials (SSVEP), lateralized readiness potentials (LRP), and frontal theta) captured with EEG in concurrent with fMRI or separate sessions. ROC analysis [123] of Distributed Equivalent Current Dipole (DECD) maps of VEP obtained from EEG in the magnet with and without active fMRI acquisition, together with DECD EEG maps obtained outside of the magnet [124] showed the effectiveness of the removal of artifacts inherent in fMRI-EEG concurrent acquisition.

4.1.1 Measuring EEG During MRI: Challenges and Approaches

Developing methods for the integrative analysis of EEG and fMRI data is difficult for several reasons, not least of which is the concurrent acquisition of EEG and fMRI itself which itself has proved challenging (see [125] for recent overview). The nature of the problem is expressed by Faraday's law of induction: a time varying magnetic field in a wire loop induces an electromotive force (EMF) proportional in strength to the area of the wire loop and to the rate of change of the magnetic field component orthogonal to the area. When EEG electrodes are placed in a strong ambient magnetic field resulting in the EMF effect, several undesirable complications arise:

- Rapidly changing MR gradient fields and RF pulses induce voltages in the EEG leads placed inside the MR scanner. Introduced potentials greatly obscure the EEG signal [126]: higher gradients magnetic field strengthen the artifacts introduced into the EEG signal [127]. This type of artifact is a real concern for concurrent EEG/MRI acquisition. Due to the deterministic nature of MR interference, hardware and algorithmic solutions may be able to unmask the EEG signal from MR disturbances. For example, Allen et al. [128] suggested an average waveform subtraction method to remove MR artifacts which is effective in case of deterministic generative process of a signal [129]. This type of method can be applied even with higher magnetic fields. A study with both phantom and animal VEP that used a 4.7 T [127], found that up to 90% of the amplitude of EEG signal acquired outside of the magnet could be restored without introducing a temporal offset. However, it is important to note that time variations of the MR artifact waveform can reduce the success of this method [91, 130]. The problem can be resolved through hardware modification that increases the precision of the synchronization of MR and EEG systems (*e.g.*, *Stepping Stone Sampling* [131]), or during post-processing by using precise timings of the MR pulses during EEG

waveform averaging [91, 132]. Algorithmic alignment of fMRI and EEG waveforms with sub-sample precision have been used successfully [133]. Acquisition of an ideal artifact waveform is also possible, via additional wire loops at the scalp. Other techniques that have been proposed to reduce MR gradient and ballistocardiographic artifacts range from technical approaches such as bipolar electrodes in a twisted configuration [134] to the use of post-processing techniques such as spectral domain filtering, spatial Laplacian filtering, PCA [135, 136, 137] Figure 4.1, ICA [see 135, 138, 139, 140, 141], and even a multistage processing pipeline [142].

- Even a slight motion of the EEG electrodes within the strong static field of the magnet can induce significant EMF [143, 144]. For instance, native pulsatile motion related to a heart beat yields a ballistocardiographic artifact in the EEG that can be roughly the same magnitude as the EEG signals themselves [126, 134]. Usually such artifacts are removed by the same average waveform subtraction, decomposition (*e.g.*, [136, 145]), multiple source correction (MSC) [146], or clustering [133] methods. One hardware solution that has been suggested for the reduction of motion effects is to use dual twist electrodes, where consecutive twists in opposite directions to the flow cancel out the effects of the motion or gradient switching [134]. To reduce the ballistocardiographic artifact electrodes are firmly bandaged to the subject's head [121].

- Induced electric currents can heat up the electrode leads to a painful, or even potentially dangerous levels [147]. Plastic electrodes with thin layers of silver chloride or gold alloys are used to provide sufficient impedance for the voltage reading on the scalp with minimal impact on the quality of the MR signal [148]. Current-limiting electric components (resistors, JFET transistors, *etc.*) are usually necessary to prevent the development of nuisance currents which can have direct

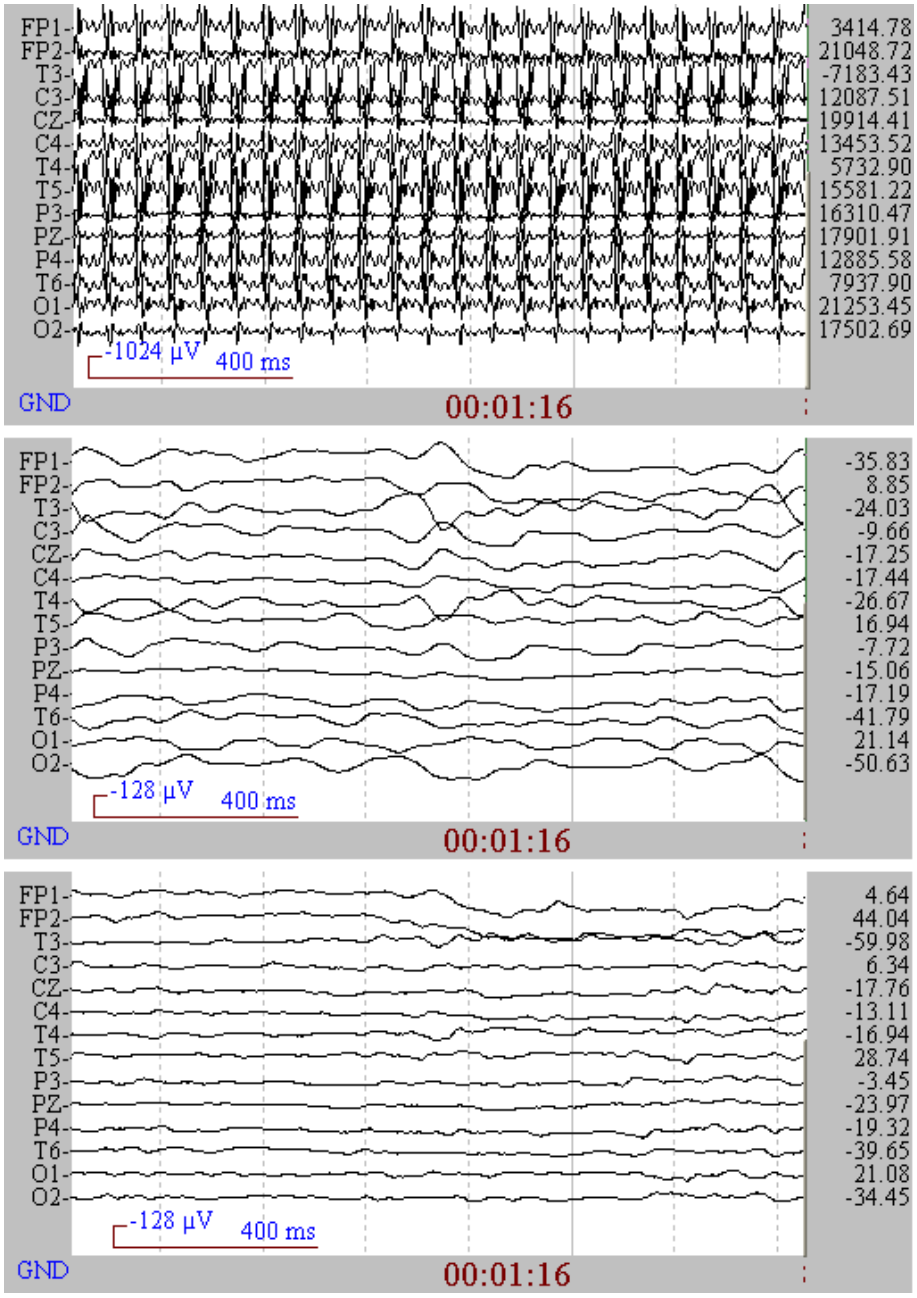


Figure 4.1 EEG MR artifact removal using PCA. EEG taken inside the magnet (top); EEG after PCA-based artifact removal but with ballistocardiographic artifacts present (center); EEG with all artifacts removed (bottom). After artifact removal it can be seen that the subject closed his eyes at time 75.9 s. (Courtesy of M. Negishi and colleagues, Yale University School of Medicine)

contact with subject's scalp. Simulations can provide the safe power range that should be used for particular configurations of coil/power/sensors in order to comply with FDA guidelines [149].

Another concern is the impact of EEG electrodes on the quality of MR images. The introduction of EEG equipment into the scanner can potentially disturb the homogeneity of the magnetic field and distort the resulting MR images [83, 126]. Recent investigations show that such artifacts can be effectively avoided [148] by using specially designed EEG equipment [134]: specialized geometries, and new "MR-safe" materials (carbon fiber, plastic) for the leads. To test the influence of a given EEG system on fMRI data, a comparison of the data collected both with and without the EEG system being present should be conducted. Analysis of data usually demonstrates the same activation patterns in two conditions [83], although a general decrease in fMRI SNR is observed when EEG is present in the magnet. A correction for the Hall effect[1] (which are used for forward E/MEG modeling) finds the following first-order correction to be negligible: $\sigma_H = 4.1 \times 10^{-8}\sigma$ for $B = 1.5\,\mathrm{T}$ [150].

4.1.2 Experimental Design Limitations

There are two ways of avoiding the difficulties associated with collecting EEG data in the magnet: (1) collect EEG and MRI data separately, or (2) use an experimental paradigm that can work around the potential contamination between the two modalities. The choice between these two alternatives will depend on the constraints associated with research goals and methodology. For example, if an experiment can be repeated more than once with a high degree of reliability of the data, separate E/MEG and fMRI acquisition may be appropriate [90, 97, 103, 105]. In cases when simultaneous measurements are essential

[1]The Hall effect is the production of a potential difference (the Hall voltage) across an electrical conductor, transverse to an electric current in the conductor and a magnetic field perpendicular to the current (http://en.wikipedia.org/wiki/Hall_effect).

for the experimental objectives (*e.g.*, cognitive experiments where a subject's state might influence the results as in monitoring of spontaneous activity or sleep state changes), one of the following protocols can be chosen:

Triggered fMRI: detected EEG activity of interest (epileptic discharge, *etc.*) triggers MRI acquisition [110, 111, 114, 151]. Due to the slowness of the HR, relevant changes in the BOLD signal can be registered 4–8 s after the event. The EEG signal can settle quickly after the end of the previous MRI block [134], so it can be acquired without artifacts caused by RF pulses or gradient fields that are present only during the MRI acquisition block. Note that ballistocardiographic and motion-caused artifacts still can be present and will require post-processing in order to be eliminated. Although this is an elegant solution and has been used with some success in the localization of epileptic seizures, this protocol does have some drawbacks. Specifically, it imposes a limitation on the amount of subsequent EEG activity that can be monitored if the EEG high-pass filters do not settle down soon after the MR sequence is terminated [106]. In this case, EEG hardware that does not have a long relaxation period must be used. Another drawback with this approach is that it requires online EEG signal monitoring to trigger the fMRI acquisition in case of spontaneous activity. Often experiments of this kind are called *EEG-correlated fMRI* due to the fact that offline fMRI data time analysis implicitly uses EEG triggers as the event onsets [129];

Interleaved EEG/fMRI: the experiment protocol consists of time blocks and only a single modality is acquired during each time-block [95, 150]. This means that every stimulus has to be presented at least once per modality. To analyze ERP and fMRI activations, the triggered fMRI protocol can be used with every stimulus presentation so that EEG and MR are sequentially acquired in order to capture a clean E/MEG signal followed by the delayed HR [85];

Simultaneous fMRI/EEG: pre-processing of the EEG signal mentioned in Section 4.1.1 is used to remove the MR-caused artifacts and to obtain an estimate of the true EEG signal. However, neither of the existing artifact removal methods has proved to be general enough to work in every type of EEG experiment and analysis. It is especially difficult to use such an acquisition scheme for cognitive experiments in which the EEG evoked responses of interest can be of small amplitude and completely overwhelmed by the MR noise [152].

4.2 Forward Modeling of BOLD Signal

Even in the case of artifact-free EEG and fMRI data acquisition, the successful analysis of data collected in a multimodal experiment remains problematic. The main problem of multimodal analysis is the absence of a general unifying account of the BOLD fMRI signal in terms of the characteristics of neuronal response. Various models have been suggested, including coarse modeling of BOLD signal in the context of a Linear Time Invariant System (LTIS) as well as general models of the BOLD signal in terms of detailed biophysical processes (MR and blood system characteristics [153], *Balloon* [154] (its linearized [155] and generalized [156] forms), or *Vein and Capillary* [157] models). The simple models are not general enough to explain the variability of the BOLD signal, whereas complex parametric models that rely heavily on prior knowledge of nuisance parameters (due to biophysical details), almost never have a reliable and straightforward means of estimation. This fact makes it unlikely that such comprehensive models can be used as reliable generative models of the BOLD signal. Currently, research is ongoing in the attempt to identify stable nuisance parameters [158] and to derive novel models suitable for data obtained in different modalities. An interesting *heuristic model* of neuronal activation and its influence on BOLD and EEG signals was recently suggested by Kilner et al. [159]. This model relates BOLD signal to

the changes in spectral characteristics of the EEG signal during activation. The proposed model formulation is consistent with the results of a number of multimodal experiments that used other forms of analysis. Thus, this model seems promising in being able to reveal reliable interdependencies between different brain imaging modalities. The following section describes modeling issues in greater detail to further underscore the limited applicability of many multimodal analysis methods covered in Section 4.3.

4.2.1 Convolutional Model of BOLD Signal

A common paradigm in early studies that collected fMRI data used simple contrast designs (*e.g.*, block design) to exploit the assumed linearity between design parameters and the HR. The belief underlying the use of block designs is that they can amplify the SNR because the HR possesses more temporal resolution than indicated by the scan acquisition time (TR).

In order to account for the present autocorrelation of the HR caused by its temporal dispersive nature, Friston et al. [160] suggested modeling HR with an LTIS, so that the HR is modeled by convolving an input (joint intrinsic and evoked neuronal activity) $q(t)$ with a hemodynamic response function (HRF) $h(t)$.

$$f(t) = (h * q)(t). \tag{4.1}$$

Because localized neuronal activity itself is not directly available through non-invasive imaging, verification of LTIS modeling on real data, as a function of parameters of the presented stimuli (*i.e.*, duration, contrast), is appropriate.

The convolutional model was used on real data to demonstrate linearity between the BOLD response and the parameters of presented stimuli [161, 162]. In fact, many experimenters have shown apparent agreement between LTIS modeling and real data, as well as its superiority over more complex, non-linear models when dealing with noisy data

[158]. Specifically, it has been possible to model responses to longer stimulus durations, by constructing them using the responses to shorter duration stimuli, and get results consistent with LTIS modeling. Because of its predictive success, its relative ease of use, and its independence from biophysical details, this modeling approach has become widely accepted. On the down side, LTIS as a modeling constraint is very weak and consequently, allows the user to make arbitrary choices of parametric HRF based solely on preference and familiarity.

Over the years multiple models for the HRF have been suggested. The most popular and widely used, until now, is a single probability density function (PDF) of the Gamma distribution by [163]. It was extended by Glover [164] to perform deconvolution of the HR signal. Nuisance parameters $(n_1, t_1, n_2, t_2, a_2)$ of the next HRF were estimated for motor and auditory areas

$$h(t) = \frac{1}{c_1}\, t^{n_1}\, e^{-t/t_1} - \frac{a_2}{c_2}\, t^{n_2}\, e^{-t/t_2} \quad \text{where} \quad c_i = \max_t t^{n_i}\, e^{-t/t_i} = \left(\frac{e}{n_i t_i}\right)^{-n_i} \tag{4.2}$$

as the sum of two unscaled PDFs of the Gamma distribution. The first term captures the positive BOLD HR and the second term captures the overshoot often observed in the BOLD signal. Many other models of HRF, both simple and sophisticated, have been suggested. These include Poisson PDF [160], Gaussians [165], Bayesian derivations [166, 167, 168], and deconvolution [169] among others. The particular choice of a given HRF model is often motivated by some factors other than those arising from bio-physics; *i.e.*, easy Fourier transformation, the presence of a post-response dip, or "best-fit" properties. Modeling of the HRF through deconvolution of the signal has been shown to improve detectability of conventional GLM analysis, for example when applied to epileptic data [119]. This is not surprising given the characteristic variability of the BOLD signal attributable to various nuisance parameters. More on this topic will be discussed in the next section (Section 4.2.2).

Following the development of the convolutional model to describe BOLD response, the issue of HR linearity became an actively debated question. If HR is linear, then with what features of the stimulus (*e.g.*, duration, intensity) or neuronal activation (*e.g.*, firing frequency, field potentials, frequency power) does it vary linearly? As a first approximation, we can define the range of possible values for those parameters in which HR was found to behave linearly. For example, early linearity tests [164] showed how difficult it was to predict long duration stimulus effects based on an estimated HR from shorter duration stimuli. [170] reviewed existing papers describing different aspects of non-linearity in BOLD HR and attempted to determine the range of values associated with linearity in three cortical areas: motor, visual and auditory complex. The results of these analyses have shown that although there is a strong non-linearity observed for small stimulus durations, long stimulus durations show a higher degree of linearity.

It appears that a simple convolutional model generally is not capable of describing the BOLD responses in terms of the experimental design parameters if such are varying in a wide range during the experiment. Nevertheless LTIS might be more appropriate to model BOLD response in terms of neuronal activation if most of the non-linearity in the experimental design can be explained by the non-linearity of the neuronal activation itself.

4.2.2 Neurophysiologic Constraints

In the previous section the subject of linearity between the experimental design parameters and the observed BOLD signal was explored. For the purpose of this work it is more relevant to explore the relation between neuronal activity and HR.

It is known that E/MEG signals are produced by large-scale synchronous neuronal activity, whereas the nature of the BOLD signal is not clearly understood. The BOLD signal does not correspond to the neural activity that consumes the most energy [171], as early researchers believed. Furthermore, the transformation between the

electrophysiological indicators of neuronal activity and the BOLD signal cannot be linear for the entire dynamic range, under all experimental conditions and across all the brain areas. Generally, a transformation function cannot be linear since the BOLD signal is driven by a number of "nuisance" physiologic processes such as cerebral metabolic oxygen consumption ($CMRO_2$), cerebral blood flow (CBF) and cerebral blood volume (CBV) as suggested by the *Balloon model* [154], which are not generally linear.

Due to the indirect nature of the BOLD signal as a tool to measure neuronal activity, in many multimodal experiments a preliminary comparative study is done first in order to assess the localization disagreement across different modalities. Spatial displacement is often found to be very consistent across multiple runs or experiments (see Section 4.3.3 for an example). Specifically, observed differences can potentially be caused by the variability in the cell types and neuronal activities producing each particular signal of interest Nunez and Silberstein [172]. That is why it is important first to discover the types of neuronal activations that are primary sources of the BOLD signal. Some progress on this issue has been made. A series of papers generated by a project to cast light on the relationship between the BOLD signal and neurophysiology, have argued that local field potentials (LFP) serve a primary role in predicting BOLD signal [173, and references 27, 29, 54, 55 and 81 therein]. This work countered the common belief that spiking activity was the source of the BOLD signal [for example 174] by demonstrating a closer relation of the observed visually evoked HR to the local field potentials (LFP) of neurons than to the spiking activity. This result places most of the reported non-linearity between experimental design and observed HR into the non-linearity of the neural response, which would benefit a multimodal analysis. Simultaneous EEG and fMRI recordings during visual stimulation [175] supported the idea that non-linearity of BOLD signal primarily reflects non-linearity of neuronal response itself by demonstrating linearity between principal current density from EEG and the estimated value of the synaptic efficacy in the modified balloon model [176].

To validate pioneering research of Logothetis et al. [177] (see Section 4.2.2) spiking activity was correlated with BOLD response across multiple sites in the cat visual cortex Kim et al. [178]. The correlation varied from point to point on the cortical surface and was generally valid only when data were averaged at least over 4-5 mm spatial scale, once again questioning spatial specificity of BOLD response.

Note that the extracellular recordings of the experiments described above, were carried out over a small ROIs, therefore they inherit the parameters of underlying hemodynamic processes for the given limited area. Thus, even if LFP is taken as the primary electrophysiological indicator of the neuronal activity causing BOLD signal, the relationship between the neuronal activity and the hemodynamic processes on a larger scale remains an open question.

Since near-infrared spectroscopy (NIRS)[2] is capable of capturing the individual characteristics of cerebral hemodynamics such as oxy-hemoglobin (HbO), deoxy-hemoglobin (HbR) and total hemoglobin (HbT) content, some researchers use NIRS to reveal the nature of the BOLD signal. Results showed great agreement between NIRS signals and BOLD, and stronger correlation between BOLD and HbR compared to HbO signals [179]. Flow response measure with ASL also had great correspondence with HbT signal.

Investigation of connection between neuronal activity and NIRS captured characteristics [180] revealed the non-linear mapping between the neuronal activity and evoked hemodynamic processes. This result should be a red flag for those who try to define the general relation between neuronal activation and BOLD signal as mostly linear. The conjoint analysis of BOLD and NIRS signals revealed the silent BOLD signal during present neural activation registered by E/MEG modalities [157]. This mismatch between E/MEG and fMRI results is known as the *sensory motor paradox* [181]. To

[2]NIRS is often used in the context of near-infrared optical imaging (NIOI), whenever multiple sensors allow to capture spatial distribution of the signal

explain this effect, the *Vein and Capillary* model was used to describe the BOLD signal in terms of hemodynamic parameters [157]. The suggested model permits the existence of silent and *negative BOLD* responses during positive neuronal activation.

The investigation of neurovascular coupling under influence of dopaminergic drugs showed that direct effects of dopamine upon the vasculature cannot be ignored [182] and can provide further insights in the nature of *negative BOLD*, which can be due to interplay between vasoconstrictive and vasodilatory substances. These facts, together with an increasing number of studies [183] suggesting that sustained negative BOLD HR is a primary indicator of decreased neuronal activation, provide yet more evidence for the inherent complexity of BOLD HR. Recently published review [184] summarized more of known factors which shape the characteristics of BOLD signal at different levels: neuronal, vascular, and signal acquisition. This paper can be referred to as the summary on available at the moment knowledge on the nature of the BOLD signal. This section concludes by noting that the absence of a generative model of the BOLD response prevents the development of universal methods of multimodal analysis. Nevertheless, as discussed in this section and is shown by the results presented in the next section, there are specific ranges of applications where the linearity between BOLD and neuronal activation can be assumed. Such simplistic model can be voted for by the supported of *Occam's razor* principle which is to prefer simple models capable of describing the data of interest.

4.3 Existing Multimodal Analysis Methods

Whenever applicable, a simple comparative analysis of the results obtained from the conventional uni-modal analyses together with findings reported elsewhere, can be considered as the first confirmatory level of a multimodal analysis. This type of analysis is very flexible, as long as the researcher knows how to interpret the results and to draw

useful conclusions, especially whenever the results of comparison reveal commonalities and differences between the two [107]. On the other hand, by default a unimodal analysis makes limited use of the data from the modalities, and encourages researchers to look for analysis methods which would incorporate the advantages of each single modality. Nevertheless, simple inspection is helpful for drawing preliminary conclusions on the plausibility to perform any conjoint analysis using one of the methods described in this section, including correlative analysis which might be considered an initial approach to try.

4.3.1 Correlative Analysis of EEG and MEG with FMRI

In some experiments, the E/MEG signal can serve as the detector of spontaneous neuronal activity (*e.g.*, epileptic discharges) or changes in the processing states (*e.g.*, vigilance states). The time onsets derived from E/MEG are alone valuable for further fMRI analysis, where the BOLD signal often cannot provide such timing information. For instance, such use of EEG data is characteristic for the experiments performed via a *Triggered fMRI* acquisition scheme (Section 4.1.2).

Correlative E/MEG/fMRI analysis becomes more intriguing if there is a stronger belief in the linear dependency between the BOLD response and features of E/MEG signal (*e.g.*, amplitudes of ERP peaks, powers of frequency components), than between the hemodynamics of the brain and the corresponding parameter of the design (*e.g.*, frequency of stimulus presentation or level of stimulus degradation). Then E/MEG/fMRI analysis effectively reduces the inherent bias present in the conventional fMRI analysis methods by removing the possible non-linearity between the design parameter and the evoked neuronal response.

The correlative analysis relies on the preprocessing of E/MEG data to extract the features of interest (ERP components such as N170 [105] or P300 [97], amplitude of

mismatch negativity (MMN) [100], contingent negative variation (CNV) [88], various characteristics of epileptic spikes [185]) to be compared with the fMRI time course with simple correlative or GLM approaches. The obtained E/MEG features first get convolved with a hypothetical HRF (Section 4.2.1) to accommodate for the HR sloppiness and are then subsampled to fit the temporal resolution of fMRI. The analysis of fMRI signal correlation with amplitudes of selected peaks of ERPs revealed sets of voxels which have a close to linear dependency between the BOLD response and amplitude of the selected ERP peak, thus providing a strong correlation ($P < 0.001$ [105]). A parametric experimental design with different noise levels introduced for the stimulus degradation [100, 105] or different levels of sound frequency deviant [100] helped to extend the range of detected ERP and fMRI activations, thus effectively increasing the significance of the results found. To support the suggested connection between the specific ERP peak and fMRI activated area, the correlation of the same BOLD signal with the other ERP peaks must be lower if any at all [105]. As a consequence, such analysis cannot prove that any specific peak of EEG is produced by the neurons located in the fMRI detected areas alone but it definitely shows that they are connected in the specific paradigm.

The search for the covariates between the BOLD signal and wide-spread neuronal signals, such as the alpha rhythm, remains a more difficult problem due to the ambiguity of the underlying process, since there are many possible generators of alpha rhythms corresponding to various functions [186] and was shown to have high variability in the rest state not only across subjects but also within a single subject [187]. As an example, Goldman et al. [92] and Laufs et al. [94] were looking for the dependency between fMRI signal and EEG alpha rhythm power during interleaved and simultaneous EEG/fMRI acquisition correspondingly. They report similar (negative correlation in parietal and frontal cortical activity), as well as contradictory (positive correlation) findings, which can be explained by the variations in the experimental setup [188] or by the heterogeneous coupling between the alpha rhythm and the BOLD response [94]. Consecutive study

[189] attempted to resolve alpha band response function (ARF) of the BOLD response. The ARF was found to differ significantly between different regions: mainly positive in the thalamus, similar in amplitude and time between occipital and left and right parietal areas, positive at the eyeball and negative at the back of the eye. It was suggested that ARF variation across different regions were due to regional differences in alpha band activity as well as due to the differences in haemodynamic response function of BOLD signal. Despite the obvious simplification of the correlative methods, they may still have a role to play in constraining and revealing the definitive forward model in multimodal applications.

4.3.2 Decomposition Techniques

The common drawback of the presented correlative analyses techniques is that they are based on the selection of the specific feature of the E/MEG signal to be correlated with the fMRI time trends, which are not so perfectly conditioned to be characterized primarily by the feature of interest. The variance of the background processes, which are present in the fMRI data and are possibly explained by the discarded information from the E/MEG data, can reduce the significance of the found correlation. That is why it was suggested [190] to use the entirety of the E/MEG signal, without focusing on its specific frequency band, to derive the E/MEG and fMRI signal components which have the strongest correlation among them. The introduction of decomposition techniques (such as basis pursuit, PCA, ICA, *etc.*) into the multimodal analysis makes this work particularly interesting.

To perform the decomposition [190], Partial Least-Squares (PLS) regression was generalized into the tri-PLS2 model, which represents the E/MEG spectrum as a linear composition of trilinear components. Each component is the product of spatial (among E/MEG sensors), spectral and temporal factors, where the temporal factors have to be maximally correlated with the corresponding temporal component of the similar fMRI

signal decomposition into bilinear components: products of the spatial and temporal factors. Analysis using tri-PLS2 modeling on the data from [92] found a decomposition into 3 components corresponding to alpha, theta and gamma bands of the EEG signal. The fMRI components found had a strong correlation only in alpha band component (Pearson correlation 0.83, $p = 0.005$), although the theta component also showed a linear correlation of 0.56, $p = 0.070$. It is interesting to note, that spectral profiles of the trilinear EEG atoms received with and without fMRI influence were almost identical, which can be explained either by the non-influential role of fMRI in tri-PLS2 decomposition of EEG, or just by a good agreement between the two. On the other hand, EEG definitely guided fMRI decomposition, so that the alpha rhythm spatial fMRI component agreed very well with the previous findings [92].

4.3.3 Equivalent Current Dipole Models

Equivalent Current Dipole (ECD) is the most elaborated and widely used technique for source localization in EMSI [see 4, for an overview of localization methods in E/MEG]. It can easily account for activation areas obtained from the fMRI analysis thus giving the necessary fine time-space resolution by minimizing the search space of non-linear optimization to the thresholded fMRI activation map. While being very attractive, such a method bears most of the problems of the ECD method (described in Chapter B), and introduces another possible bias due to the belief in the strong coupling between hemodynamic and electrophysiological activities. For this reason it needs to be approached with caution in order to carefully select the fMRI regions to be used in the ECD/fMRI combined analysis.

Although good correspondence between ECD and fMRI results is often found [191], some studies reported a significant (1–5 cm) displacement between locations obtained from fMRI analysis and ECD modeling [89, 116, 192, 193, 194]. The

displacement often was found to be very consistent across the experiments of different researchers using the same paradigm (for instance motor activations [89, 90, 195]). To assess the importance of found displacements it is necessary to keep in mind proven lower bounds on localization errors adherent for the cases of precise head modeling and mis-localization due to nuisance parameters (*e.g.*, number and models of sources, tissue conductivities and their anisotropy, head geometry) misspecification which can drastically degrade localization performance.

As it was already mentioned, in the first step, a simple comparison of detected activations across the two modalities can be done to increase the reliability of dipole localization alone. Further, additional weighting by the distance from the ECD to the corresponding fMRI activation foci can guide ECD optimization [196] and silent in fMRI activations can be accommodated by introducing free dipoles without the constraint on dipole location. Auxiliary fMRI results can help to resolve the ambiguity of the inverse E/MEG problem if ECD lies in the neighborhood of multiple fMRI activations. Placing multiple ECDs inside the fMRI foci with successive optimization of ECDs orientations and magnitudes may produce more meaningful results, especially if it better describes the E/MEG signal by the suggested multiple ECDs model.

Due the large number of consistent published fMRI results [197], it seems viable to perform a pure E/MEG experiment with consequent ECD analysis using known relevant fMRI activation areas found by the other researchers performing the same kind of experiment [198], thus providing the missing temporal explanation to the known fMRI activations.

4.3.4 Linear Inverse Methods

Dale and Sereno [199] formulated a simple but powerful linear framework for the integration of different imaging modalities into the inverse solution of DECD, where the

solution was presented as unregularized (just minimum-norm) (B.7) with $\mathbf{W_Q} = \mathbf{C}_S$ and $\lambda \mathbf{W_X} = \mathbf{C}_\epsilon$.

The simplest way to account for fMRI data is to use thresholded fMRI activation map as the inverse solution space but this was rejected [200] due to its incapability to account for fMRI silent sources, which is why the idea to incorporate variance information from fMRI into $\mathbf{C}_S$ was further elaborated [201] by the introduction of relative weighting for fMRI activated voxels via constructing a diagonal matrix $\mathbf{W_Q} = \mathbf{W}_{\text{fMRI}} = \{\nu_{ii}\}$, where $\nu_{ii} = 1$ for fMRI activated voxels and $\nu_{ii} = \nu_0 \in [0, 1]$ for voxels which are not revealed by fMRI analysis. A Monte Carlo simulation showed that $\nu_0 = 0.1$ (which corresponds to the 90% relative fMRI weighting) leads to a good compromise with the ability to find activation in the areas which are not found active by fMRI analysis and to detect active fMRI spots (even superficial) in the DECD inverse solution. An alternative formulation of the relative fMRI weighting in the DECD solution can be given using a subspace regularization (SSR) technique [202], in which an E/MEG source estimate is chosen from all possible solutions describing the E/MEG signal, and is such that it minimizes the distance to a subspace defined by the fMRI data (see Figure 4.2). Such formulation helps to understand the mechanism of fMRI influence on the inverse E/MEG solution: SSR biases underdetermined the E/MEG source locations toward the fMRI foci.

The relative fMRI weighting was tested [203] in an MEG experiment and found conjoint fMRI/MEG analysis results similar to the results reported in previous fMRI, PET, MEG and intracranial EEG studies. Babiloni et al. [204] followed Dale et al. [203] in a high resolution EEG and fMRI study to incorporate non-thresholded fMRI activation maps with other factors. First of all, the $\mathbf{W}_{\text{fMRI}}$ was reformulated to $(\mathbf{W}_{\text{fMRI}\prime})_{ii} = \nu_0 + (1 - \nu_0)\Delta_i/\Delta_{\text{max}}$, where Δ_i corresponds to the relative change of the fMRI signal in the i-th voxel, and Δ_{max} is the maximal detected change. This way the relative E/MEG/fMRI scheme is preserved and locations of stronger fMRI activations have higher prior variance. Finally the three available weighting factors were combined:

fMRI relative weighting, correlation structure obtained from fMRI described by the matrix of correlation coefficients $\mathbf{K}_S$, and the gain normalization weighting matrix $\mathbf{W}_{\text{n}}$: $\mathbf{W}_{\mathbf{Q}} = \mathbf{W}_{\text{fMRI}\prime}^{1/2}\mathbf{W}_{\text{n}}^{1/2}\mathbf{K}_S\mathbf{W}_{\text{n}}^{1/2}\mathbf{W}_{\text{fMRI}\prime}^{1/2}$. Although $\mathbf{W}_{\text{fMRI}\prime}$ alone had improved EMSI localization, the incorporation of the $\mathbf{K}_S$ lead to finer localization of neuronal activation associated with finger movement.

Although most of the previously discussed DECD methods are involved in finding minimal L_2 norm solution, the fMRI conditioned solution with minimal L_1 norm (regularization term in (B.5) $\mathcal{C}(\mathbf{Q}) = \|\mathbf{Q}\|_1$) is shown to provide a sparser activation map [205] with activity focalized to the seeded hotspot locations [196].

An fMRI-conditioned linear inverse is an appealing method due to its simplicity, and rich background of DECD linear inverse methods derived for the analysis of E/MEG signals. Nonetheless, one should approach these methods with extreme caution in a domain where non-linear coupling between BOLD and neural activity is likely to overwhelm any linear approximation [193].

4.3.5 Beamforming

Lahaye et al. [206] suggest an iterative algorithm for conjoint analysis of EEG and fMRI data acquired simultaneously during an event-related experiment. Their method relies on iterated source localization by the linearly constrained minimum variance (LCMV) beamformer (B.9), which makes use of both EEG and fMRI data. The covariance $\mathbf{C}_X$ used by the beamformer is calculated anew each time step, using the previously estimated sources and current event responses from both modalities. This way neuronal sites with a good agreement between the BOLD response and EEG beamformer reconstructed source amplitude, benefit most at each iteration. Although the original formulation is cumbersome, this method appears promising as (a) it makes use of both spatial and temporal information available from both modalities, and (b) it can account

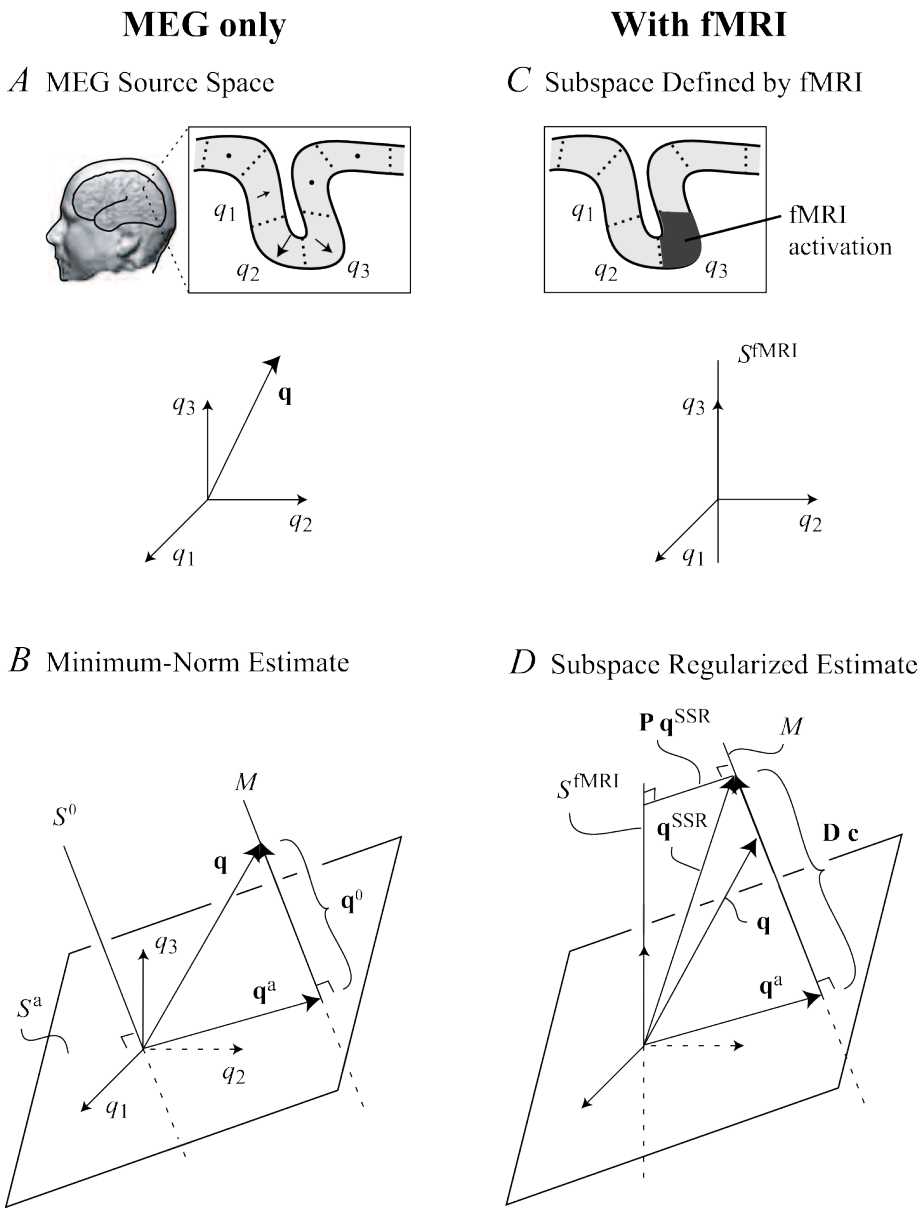


Figure 4.2 Geometrical interpretation of subspace regularization in the MEG/EEG source space. (A) The cerebral cortex is divided into source elements $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_K$, each representing an ECD with a fixed orientation. All source distributions compose a vector $\mathbf{q}$ in K-dimensional space. (B) The source distribution $\mathbf{q}$ is divided into two components $\mathbf{q}^a \in S^{\text{a}} \equiv \text{range}(\mathbf{G}^\top)$, determined by the sensitivity of MEG sensors and $\mathbf{q}^0 \in \text{null}\,\mathbf{G}$, which does not produce an MEG signal. (C) The fMRI activations define another subspace S^{fMRI}. (D) The subspace-regularized fMRI-guided solution $\mathbf{q}^{\text{SSR}} \in \mathbf{M}$ is closest to S^{fMRI}, minimizing the distance $\|\mathbf{P}\mathbf{q}^{\text{SSR}}\|$, where $\mathbf{P}$ (a $N \times N$ diagonal matrix with $\mathbf{P}_{ii} = 1/0$ when the i-th fMRI voxel is active/inactive) is the projection matrix into the orthogonal complement of S^{fMRI}. (Adapted from [202, Figure 1])

for silent BOLD sources using an electro-metabolic coupling constant which is estimated for each dipole and defines the influence of the BOLD signal at a given location onto the estimation of $\mathbf{C}_S$ which, in turn, drives the estimate of $\mathbf{C}_X$.

4.3.6 Bayesian Inference

During the last decade, Bayesian methods became dominant in the probabilistic signal analysis. The idea behind them is to use Bayes' rule to derive a *posterior probability* of a given *hypothesis* having observed data $\mathcal{D}$, which serves as *evidence* to support the hypothesis

$$p(\mathcal{H}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathcal{H})\, p(\mathcal{H})}{p(\mathcal{D})}, \tag{4.3}$$

where $p(\mathcal{H})$ and $p(\mathcal{D})$ are prior probabilities of the hypothesis and evidence correspondingly, and the conditional probability $p(\mathcal{D}|\mathcal{H})$ is known as a *likelihood function*. Thus, (4.3) can be viewed as a method to combine the results of conventional likelihood analyses for multiple hypotheses into the posterior probability of the hypotheses $p(\mathcal{H}|\mathcal{D})$ or some function of it, after been exposed to the data. The derived posterior probability can be used to select the most probable hypothesis, *i.e.*, the one with the highest probability

$$\hat{\mathcal{H}}_{|\mathcal{D}} = \arg\max_{\mathcal{H}} p(\mathcal{H}|\mathcal{D}) = \arg\max_{\mathcal{H}} \log p(\mathcal{D}|\mathcal{H}) + \log p(\mathcal{H}) \tag{4.4}$$

leading to the maximum *a posteriori* (MAP) estimate, where the prior data probability $p(\mathcal{D})$ (often called a *partition function*) is omitted because the data does not depend on the choice of the hypothesis and it does not influence the maximization over $\mathcal{H}$.

For the class of problems related to the signal processing, hypothesis $\mathcal{H}$ generally consists of a model $\mathcal{M}$ characterized by a set of nuisance parameters $\mathbf{\Theta} = \{\theta_1, \theta_{2\ldots n}\}$. The primary goal usually is to find a MAP estimate of some quantity of interest $\mathbf{\Delta}$ or, more generally, its posterior probability distribution $p(\mathbf{\Delta}|\mathcal{D}, \mathcal{M}, \mathbf{\Theta})$. $\mathbf{\Delta}$ can be an

arbitrary function of the hypothesis or its components $\boldsymbol{\Delta} = f(\mathcal{H})$, or often just a specific nuisance parameter of the model $\boldsymbol{\Delta} \equiv \theta_1$. To obtain posterior probability of the nuisance parameter, its marginal probability has to be computed by the integration over the rest of the parameters of the model

$$p(\theta_1|\mathcal{D},\mathcal{M}) = \int p(\theta_1,\theta_{2\ldots n}|\mathcal{D},\mathcal{M})\, d\theta_{2\ldots n} = \int p(\theta_1|\theta_{2\ldots n},\mathcal{D},\mathcal{M})\, p(\theta_{2\ldots n}|\mathcal{D},\mathcal{M})\, d\theta_{2\ldots n}. \tag{4.5}$$

Due to the integration operation involved in determination of any marginal probability, Bayesian analysis becomes very computationally intensive if analytical integral solution does not exist. Therefore, sampling techniques (*e.g.*, MCMC, Gibbs sampler) are often used to estimate full posterior probability $p(\boldsymbol{\Delta}|\mathcal{D},\mathcal{M})$, MAP $\hat{\boldsymbol{\Delta}}_{|\mathcal{D},\mathcal{M}} = \arg\max_{\boldsymbol{\Delta}} p(\boldsymbol{\Delta}|\mathcal{D},\mathcal{M})$, or some statistics such as an expected value $E[\boldsymbol{\Delta}|\mathcal{D},\mathcal{M}]$ of the quantity of interest.

The Bayesian approach sounds very appealing for the development of multimodal methods. It is inherently able to incorporate all available evidence, which is in our case obtained from the fMRI and E/MEG data ($\mathcal{D} = \{\mathbf{X},\mathbf{F}\}$) to support the hypothesis on the location of neuronal activations, which is in the case of DECD model is $\mathcal{H} = \{\mathbf{Q},\mathcal{M}\}$. However, the detailed analysis of (4.3) leads to necessary simplifications and assumptions of the prior probabilities in order to derive a computationally tractable formulation. Therefore it often loses its generality. Thus to derive a MAP estimator for $\hat{\mathbf{Q}}_{|\mathbf{X},\mathbf{B},\mathcal{M}}$ Trujillo-Barreto et al. [207] had to condition the computation by a set of simplifying modeling assumptions such as: noise is normally distributed, nuisance parameters of forward models have inverse Gamma prior distributions, and neuronal activation is described by a linear function of hemodynamic response. The results on simulated and experimental data from a somatosensory MEG/fMRI experiment confirmed the applicability of Bayesian formalism to the multimodal imaging even under the set of simplifying assumptions mentioned above.

Usually, model $\mathcal{M}$ is not explicitly mentioned in Bayesian formulations (such as (4.5)) because only a single model is considered. For instance, Bayesian formulation of LORETA E/MEG inverse corresponds to a DECD model, where $\boldsymbol{\Theta} = \mathbf{Q}$ is constrained to be smooth (in space), and to cover whole cortex surface. In the case of the *Bayesian Model Averaging* (BMA), the analysis is carried out for different models $\mathcal{M}_i$, which might have different nuisance parameters, *e.g.*, E/MEG and BOLD signals forward models, possible spatial locations of the activations, constraints to regularize E/MEG inverse solution. In BMA analysis we combine results obtained using all considered models to compute the posterior distribution of the quantity of interest

$$p(\boldsymbol{\Delta}|\mathcal{D}) = \sum_i p(\boldsymbol{\Delta}|\mathcal{D}, \mathcal{M}_i)\, p(\mathcal{M}_i|\mathcal{D}), \tag{4.6}$$

where the posterior probability $p(\mathcal{M}_i|\mathcal{D})$ of any given model $\mathcal{M}_i$ is computed via Bayes' rule using prior probabilities $p(\mathcal{M}_i)$, $p(\mathcal{D})$ and the likelihood of the data given each model

$$p(\mathcal{D}|\mathcal{M}_i) = \int p(\mathcal{D}|\boldsymbol{\Theta}, \mathcal{M}_i)\, p(\boldsymbol{\Theta}|\mathcal{M}_i)\, d\boldsymbol{\Theta}. \tag{4.7}$$

Initially, BMA was introduced into the E/MEG imaging [208], where Bayesian interpretation of (B.7) was formulated to obtain $p(\mathbf{Q}|\mathbf{X}, \mathbf{F})$ for the case of Gaussian uncorrelated noise ($\mathbf{W}_\mathbf{X} = \mathbf{C}_\epsilon = \nu_\epsilon \mathbf{I}$). In order to create a model, the brain volume gets partitioned into a limited set of spatially distinct functional compartments, which are arbitrarily combined to define a $\mathcal{M}_i$, search space for the E/MEG inverse problem.

At the end, different models are sampled from the posterior probability $p(\mathcal{M}_i|\mathbf{X})$ to get the estimate of the expected activity distribution of ECDs over all considered source models

$$E[\mathbf{Q}|\mathbf{X}] = \sum_i E[\mathbf{Q}|\mathbf{X}, \mathcal{M}_i]\, p(\mathcal{M}_i|\mathbf{X})$$

$$\mathrm{Var}[\mathbf{Q}|\mathbf{X}] = \sum_i \mathrm{Var}[\mathbf{Q}|\mathbf{X}, \mathcal{M}_i]\, p(\mathcal{M}_i|\mathbf{X}),$$

where the normalized probability $p(\mathcal{M}_i|\mathbf{X})$, Bayes' Factor B_{i0}, and prior odds α_i, are

$$p(\mathcal{M}_i|\mathbf{X}) = \frac{\alpha_i B_{i0}}{\sum_k \alpha_k B_{k0}} \qquad B_{i0} = \frac{p(\mathbf{X}|\mathcal{M}_i)}{p(\mathbf{X}|\mathcal{M}_0)} \qquad \alpha_i = \frac{p(\mathcal{M}_i)}{p(\mathcal{M}_0)}$$

In the original BMA framework for E/MEG [208] $\alpha_i = 1\,\forall i$, *i.e.*, the models had a flat prior PDF because no additional functional information was available at that point. Melie-García et al. [209] suggested to use the significance values of fMRI statistical t-maps to derive $p(\mathcal{M}_i)$ as the mean of all such significance probabilities across the present in $\mathcal{M}_i$ compartments. This strategy causes the models consisting of the compartments with significantly activated voxels get higher prior probabilities in BMA. The introduction of fMRI information as the prior to BMA analysis reduced the ambiguity of the inverse solution, thus leading to better localization performance. Although further analysis is necessary to define the applicability range of the BMA in E/MEG/fMRI fusion, it already looks promising because of the use of fMRI information as an additional evidence factor in E/MEG localization rather than a hard constraint.

Due to the flexibility of Bayesian formalism, various Bayesian methods solving E/MEG inverse problem already can be easily extended to partially accommodate evidence obtained from the analysis of fMRI data. For instance, correlation among different areas obtained from fMRI data analysis can be used as a prior in the Bayesian reconstruction of correlated sources [210]. The development of a neurophysiologic generative model of BOLD signal would allow many Bayesian inference methods (such as [211]) to introduce complete temporal and spatial fMRI information into the analysis of E/MEG data.

4.4 Suggested Multimodal Analysis Method

The only reason some people get lost in thought is because it's unfamiliar territory

– Paul Fix

4.4.1 Motivation

As it was shown in the previous sections, fMRI BOLD signal is an inherently non-linear function of neuronal activation. Nevertheless there have been multiple reports of linear dependency between the observed BOLD response and the selected set of E/MEG signal features. In general, such results are not inconsistent with the non-linearity of BOLD, since a non-linear function can be approximately linear in the context of a specific experimental design, regions of interest, or dynamic ranges of the selected features of E/MEG signals. Besides dominant LFP/BOLD linearity reported by Logothetis and Wandell [173] and also confirmed in the specific frequency bands of EEG signal during flashing checkerboard experiment [212], there have been reports of a strong correlation between the BOLD and the power of different frequency bands of E/MEG signals [213]. Besides conventionally explored frequency bands of E/MEG, very low-frequencies ($f < 0.1\,\mathrm{Hz}$, also known as DC-E/MEG) signal component has not yet been a subject of attention for multimodal integration despite recent experiments showing the strong correlation between the changes of the observed DC-EEG signal and the hemodynamic changes in the human brain [214]. In fact, such DC-E/MEG/BOLD coupling suggests that the integration of fMRI and DC-E/MEG might be a particularly useful way to study the nature of the time variations in HR signal. These variations are usually observed during fMRI experiments but are not explicitly explained by the experimental design or by the physics of MR acquisition process.

To summarize, independent researchers have detected various features of the E/MEG signals which have strong correspondence to the BOLD signal. Since the spectral power of E/MEG signal at different frequency bands was reported to correlate significantly with the BOLD response, it is worthwhile to approach the *fusion* problem by looking for a way to describe BOLD signal in terms of the time-varying spectral representation of E/MEG.

4.4.2 Formulation of the Approach

The overall goal of the presented analysis approach is to determine the mapping function $\mathcal{F}$, such that

$$\mathbf{F}_i = \mathcal{F}_i(\mathbf{X}), \tag{4.8}$$

where fMRI signal $\mathbf{F}$ at each voxel i is described in terms of the E/MEG data $\mathbf{X}$ via a transformation $\mathcal{F}_i$.

Due to aforementioned loose or absent coupling between power of some frequency bands and the design of the experiment, it is desirable to analyze the data from both modalities, whenever they are acquired in a single session. Due to impossibility of acquiring MEG in the magnet, suggested method should mostly be applicable to the conjoint analysis of EEG and fMRI. Suggested approach is similar to GLM analysis of fMRI data where each voxel is represented as a linear combination of a limited number of explanatory variables. However, it substantially differs from GLM analysis since it neither imposes a specific shape of HRF, nor, strictly speaking, it requires linear relationship between $\mathbf{X}$ and $\mathbf{F}_i$. Furthermore, it does not require the specification of the experimental design.

Richness of E/MEG signal (*e.g.*, in terms of its spectral properties), absence of a generative forward model of BOLD signal, and sluggishness of BOLD response are the main obstacles toward deriving a reliable transformation function $\mathcal{F}$. Slow evolution of

BOLD response requires to rely on a vast duration of E/MEG signal preceding any given volume of a BOLD signal. Therefore, to accommodate for the existing lag between neural activation and its reflection within fMRI signal, (4.8) can be further elaborated as

$$\mathbf{F}_i(t) = \mathcal{F}_i(\mathbf{X}(t-\tau \ldots t)), \tag{4.9}$$

where t is a moment in time when an fMRI volume is acquired (fMRI volumes are usually evenly spaced in time with a fixed TR, typically 2–4 sec), and τ is a reasonable duration to look back in time for neural activations which should be significantly reflected in BOLD response at the given time point. Such setup is schematically depicted in Figure 4.3. In most of the scenarios, it is reasonable to take $\tau = 10 +$ TR sec. The duration of 10 sec should be sufficient to account for the typical HRF delay. The duration of TR is included since different slices of a volume are acquired at different offsets within the TR. Hence for atypically long TRs (as it is in the experimental data of Section 4.6, where TR=10 sec) it is necessary to consider sufficient amount of E/MEG data to describe any possible voxel in the acquired volume.

It is possible to account for the offset of any given slice while formulating (4.9), if slice ordering is known in advance. But such procedure might be inappropriate since fMRI data are conventionally motion corrected as a part of preprocessing. Motion correction relies on the interpolation between different slices, hence it can *smear* the effect of slice acquisition timing. Furthermore, transformation $\mathcal{F}$ could be described as a composition of actual regression function $\mathcal{G}$ and some (optional) time-frequency transformation $\mathcal{T}$ (*e.g.*, SFFT, Wavelet standard or packet decomposition)

$$\mathbf{F}_i(t) = \mathcal{G}_i(\mathcal{T}(\mathbf{X}(t-\tau \ldots t))), \tag{4.10}$$

where $\mathcal{T}$ does not depend on the voxel in question and is introduced to convert the original E/MEG data into the feature space which is known to have correspondence with BOLD response. Transformation $\mathcal{T}$ can also incorporate multiple characteristics of the

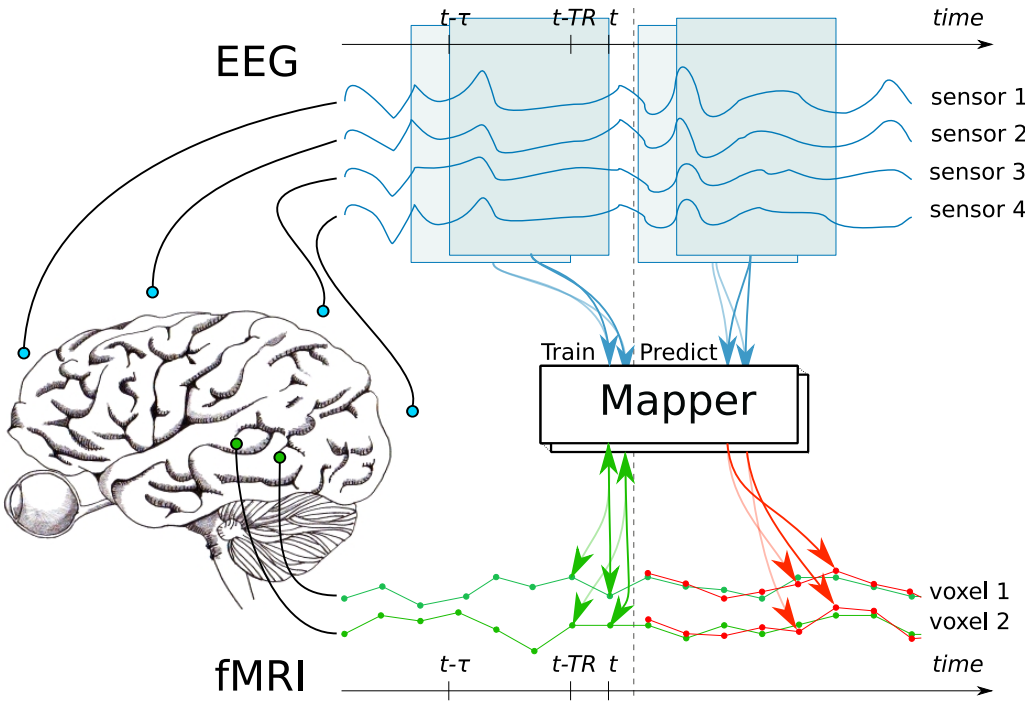


Figure 4.3 Mapping of neural data from EEG into fMRI. During learning, for each fMRI voxel a single mapper is trained on a corresponding time-window of EEG data to predict fMRI signal in that voxel. During testing the trained mapper predicts fMRI signal. Comparison between predicted (in red) and real time course of fMRI (in green) allows to assess the performance of the mapper.

same feature of the E/MEG signal, *e.g.*, for a single time-frequency slot it can provide both amplitude and the power of the E/MEG signal. That can allow to account for some non-linear effects which were reported for the BOLD response without requiring transformation $\mathcal{G}$ itself to be non-linear. Such approach is substantially different from the existing multimodal methods. None of the existing methods (such as correlative analysis Section 4.3.1) presented a reliable and generative (defining the actual transformation $\mathcal{F}$) way to describe BOLD signal in terms of multiple components of E/MEG signal. Furthermore, they often had to rely on some chosen, hence rigid, HRF function. That heavily restricts the amount of novel conclusions to be drawn from the data analysis.

Having selected features of the signals to be involved in the fusion, many EMSI methods could naturally be extended to account for fMRI data if a generative forward model of BOLD signal was available. For instance, direct universal-approximator inverse methods [215, 216] have been found to be very effective (fast, robust to noise and to complex forward models) for the E/MEG dipole localization problem, and could be augmented to accept fMRI data if the generative model for it was provided. Due to the mentioned forward model of BOLD and abundance of the features of E/MEG signal, instead of relying on a overly too flexible generic architecture for the description of the signal (*e.g.*, artificial neural networks as in [215, 216]), it is logical to use some other supervised machine learning regression approaches (*e.g.*, SVR and GPR) which are inherently well regularized, and as a result are capable to provide reasonable generalization performance even if the dimensionality of input space greatly exceeds the number of available data samples.

4.4.3 Promises

As it was shown in Chapter 3, methods such as GPR and SVM are capable of near-perfect generalization performance on classification tasks, even when

dimensionality of input space greatly exceeded number of data samples. Furthermore, supervised learning methods encourage model testing to provide unbiased estimates of the generalization performance for a derived transformation $\mathcal{F}$. At last, many methods (*e.g.*, linear ones) conveniently provide feature-weighting, which can be used for determining the components of the signal relevant for a given classification or regression task.

As it will be shown in the followup sections, derivation of a reliable transformation $\mathcal{F}$ can help to

- identify brain regions, where fMRI signal is relatively well explained by the information present in E/MEG signal. It provides insights not only about the nature of the fMRI response, but also about interconnectivity among the areas, since deep structures themselves are weakly represented in the E/MEG signal, hence goodness of their description in terms of the E/MEG signal relies on their connections with more exterior parts of the brain;

- localize the regions which are related to specific components of E/MEG signal (*e.g.*, different frequency bands);

- provide sensitivity maps across the regions for the E/MEG per each channel, which in turn would allow to characterize coherence among distant parts of the brain due to inherent connectivity;

- filter E/MEG signal based on fMRI data, or even to perform electrode rejection based on its participation in the description of fMRI data;

- filter, interpolate, or predict (thus increase temporal sampling) of fMRI signal based on the E/MEG signal;

- assess spatial evolution of HRF of BOLD fMRI response per each voxel;

4.4.4 Computational Efficiency

Since the suggested method proceeds in a mass-univariate fashion by estimating transformation function $\mathcal{G}_i$ per each target voxel, it sounds like an idea destined for a failure since typical fMRI volume has up to 200,000 brain voxels (actual number of voxels depends on the spatial resolution of the data). However, in the suggested approach each input sample for the regression is the set of features of E/MEG data, which are constant across all voxels. The time-course of fMRI voxel is the only varying variable, different per each voxel. Therefore, most of the kernel based ML methods, for a specific set of parameters, need to compute the kernel matrices only once. Initial optimization point for the iterative ML methods (such as constrained quadratic optimization in SVM) can be pre-seeded with previously found solution for another voxel. Due to guaranteed uniqueness of the solution, such initialization of optimization would result in faster convergence, and must result in nearly exact (up to a specified numerical tolerance) solution as if it was done with full training and arbitrary initial starting point for optimization. That allows to perform training of the regression per each voxel in a reasonable amount of time (*e.g.*, from ten to hundreds of milliseconds), therefore making whole brain analysis feasible. Due to the mass-univariate nature, estimation can also be parallelized for the computation in the high performance computing environments.

For the analysis presented here, SVR was chosen due to its simple parametrization (just the trade-off coefficient C with a reasonable default) and computational efficiency. Comparison of the results obtained with different regression methods is left for future research.

4.5 Validation on a Simulated Data

As previously emphasized, any novel methodology has to be validated first on data with the known characteristics of the noise and of the signal of interest. Due to the absence of a

realistic study of the phantom, it is necessary to simulate the signal and noise conditions. This chapter describes the protocol used to simulate the dataset and presents the results of the analysis using the suggested methodology.

4.5.1 Simulation

Simulated environment consisted of a single active slice within a three shell (brain, skull, scalp) spherical model of the head (see Figure 4.4). FMRI resolution was taken to be isotropic 3 mm. At this resolution the modeled "brain" consisted of 2148 voxels. Temporal resolutions for the simulated signals were taken to be in line with the ones of real dataset (kindly provided by Dr.Herrmann [217], analysis of which will be presented in Section 4.6): 50 Hz for EEG, and 0.2 Hz (equivalent to TR=5 sec) for fMRI. Three regions of interest (ROIs), colored in red, dark red, and blue in Figure 4.4, were defined to carry event related (ER) activity, whereas all other locations carried only spontaneous activity, amount of which depended on the type of the matter (gray or white). For EEG modeling, each voxel contained a single dipole, orientation of which (red arrows in Figure 4.4) was set along the normal to the artificial folding (depicted in yellow) of the cortical tissue. Eight out of 16 EEG sensors (labeled as S0...S7) were located in the plane of fMRI slice. Other 7 (S8...S14) were located in the slice half-way to the top of the head, and the last channel S15 was located on top of the head. Forward model for EEG signal was estimated using *OpenMEEG*[3] (see Figure C.3).

Two types of neural activations were modeled – event-related (ER) and spurious activity. Event-related activity was added only within predefined ROIs. Onsets of ER activity were taken from the experimental design in the real dataset of Section 4.6. Relative amplitudes in the ROI1...ROI3 (see Figure 4.4) of the activations were in the range from 0.5 to 1.3 depending on the power of auditory stimulation (60 dB vs 80 dB)

[3] http://www-sop.inria.fr/odyssee/software/OpenMEEG

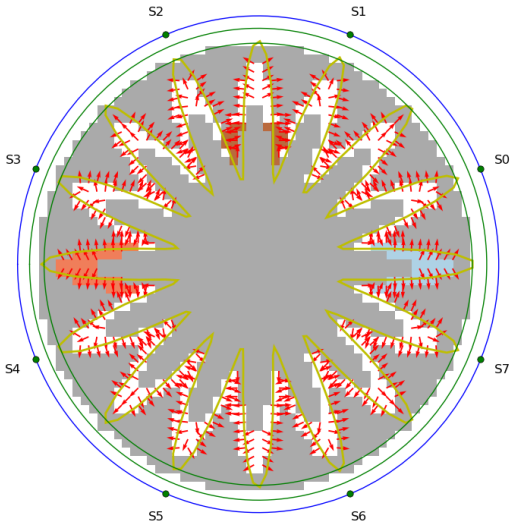
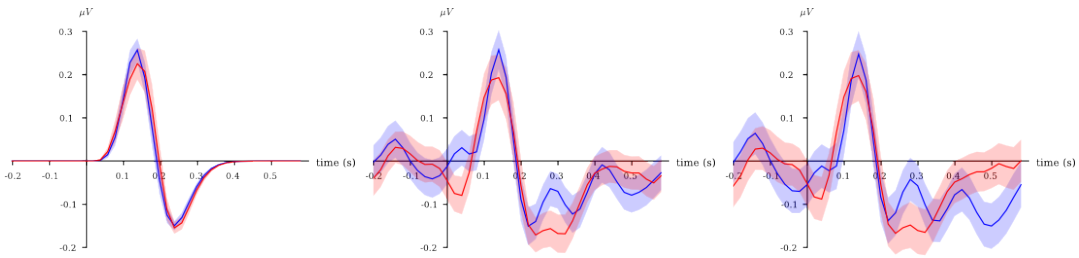


Figure 4.4 Simulated Environment Space. 2148 voxels in a single plane. 3 ROIs were defined to carry ER activity. Arrows (in red) visualize the orientation of dipoles within gray matter voxels. 8 out of 16 EEG sensors are located on the scalp surface.

and the laterality of the area voxel belonged to in respect to the side of auditory stimulation (left vs right).

Since various parts of the brain are active all the time, there always exists neural activity which is not related to the activity caused by the external stimulation, or the experimental design in general. Nevertheless, such spurious activity effects both EEG and fMRI responses but does not correlate with the design, thus it cannot be accounted for in GLM analysis of fMRI or in the ERP analysis of EEG. Hence, this activity contributes variance to the acquired neural data which cannot be explained by the variables of the experimental design, and it is different from the instrumental noise which should not have common cause in different data modalities. For the simulation, a range of different types of spurious activity were assigned randomly to different locations (see Figure C.1 for examples of the shapes of ERP responses). Amplitude and frequency of occurrence of spurious activity varied depending on the type of the simulated matter of the brain. Within



(a) Event-related (b) Total (ER and spurious) (c) Acquired

Figure 4.5 ERP of the simulated EEG data.

gray matter voxels spurious activity occurred with a frequency of 2 events per second and relative amplitude of 0.6. Within white matter voxels frequency and amplitude were half a small.

While modeling EEG responses, all simulated neural activity was jitterred uniformly within 50 msec of stimuli appearance to simulate the variability of the response per trial. Additionally, all locations in the "brain" were assigned an additional jitter (once again random within 50 msec) to model the variability of the response due to variability of the activation depending on the location. Both types of activations contributed to the total local field potential (LFP) signal assigned to each voxel. LFP signals of all of the brain voxels were transformed into EEG signals through the aforementioned forward EEG model. Jitterring and addition of spontaneous activity resulted in ERP responses, which looked very similar to the ones acquired empirically (see Figure 4.5(b)). Additional additive source of variance, instrumental noise, was modeled by random sampling of normal distribution with mean 0 and variance 1, with a consecutive low-pass Butterworth filtering in temporal domain with stop-band frequency of 4 Hz (see Figure 4.6).

Simple linear convolutional model (see Section 4.2.1) of fMRI signal was chosen. Squared values of the activations amplitudes were convolved with a conventional HRF (Figure C.2). Additive instrumental noise was modeled in the fashion similar to EEG

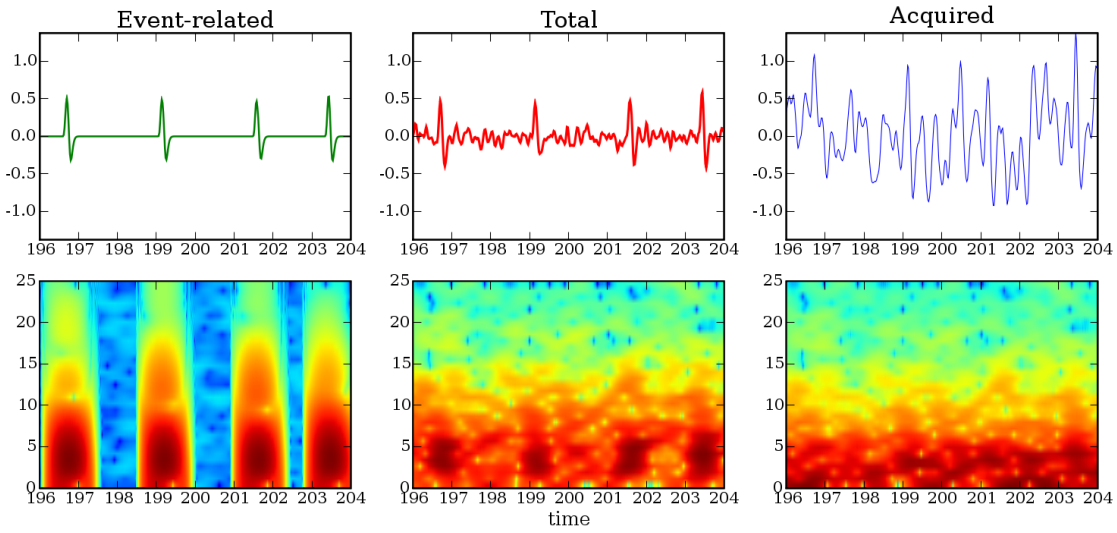


Figure 4.6 Spectral characteristics of simulated EEG signals. The upper part presents sample durations of EEG. The lower part shows corresponding spectrograms.

noise, but with a different stop-band frequency of the filter at 0.2 Hz. Samples of the modeled signals are presented on the Figure 4.7.

4.5.2 Data Preparation

EEG recordings of brain activity are dominated by the periods of rhythmic activity. It was shown [218] that characteristic bands of EEG signal, as generated by the neural matter, are centered at frequencies arranged according to the powers of $\phi = 1.61803399$ (the *"golden ratio"*[4]) (see Figure 4.8).

Thus, prior the analysis, EEG signal of each channel was transformed into its power spectra, and was summarized by the total power in each temporal/spectral bin taking $\tau = 10$ sec (see (4.10)) seconds prior the onset of each fMRI volume acquisition. Temporal width of each bin was taken to be 500 ms and frequencies followed aforementioned ϕ distance in a log space to cover the range up to 25 Hz available within simulated signal sampled at 50 Hz. That resulted in

[4] http://en.wikipedia.org/wiki/Golden_ratio

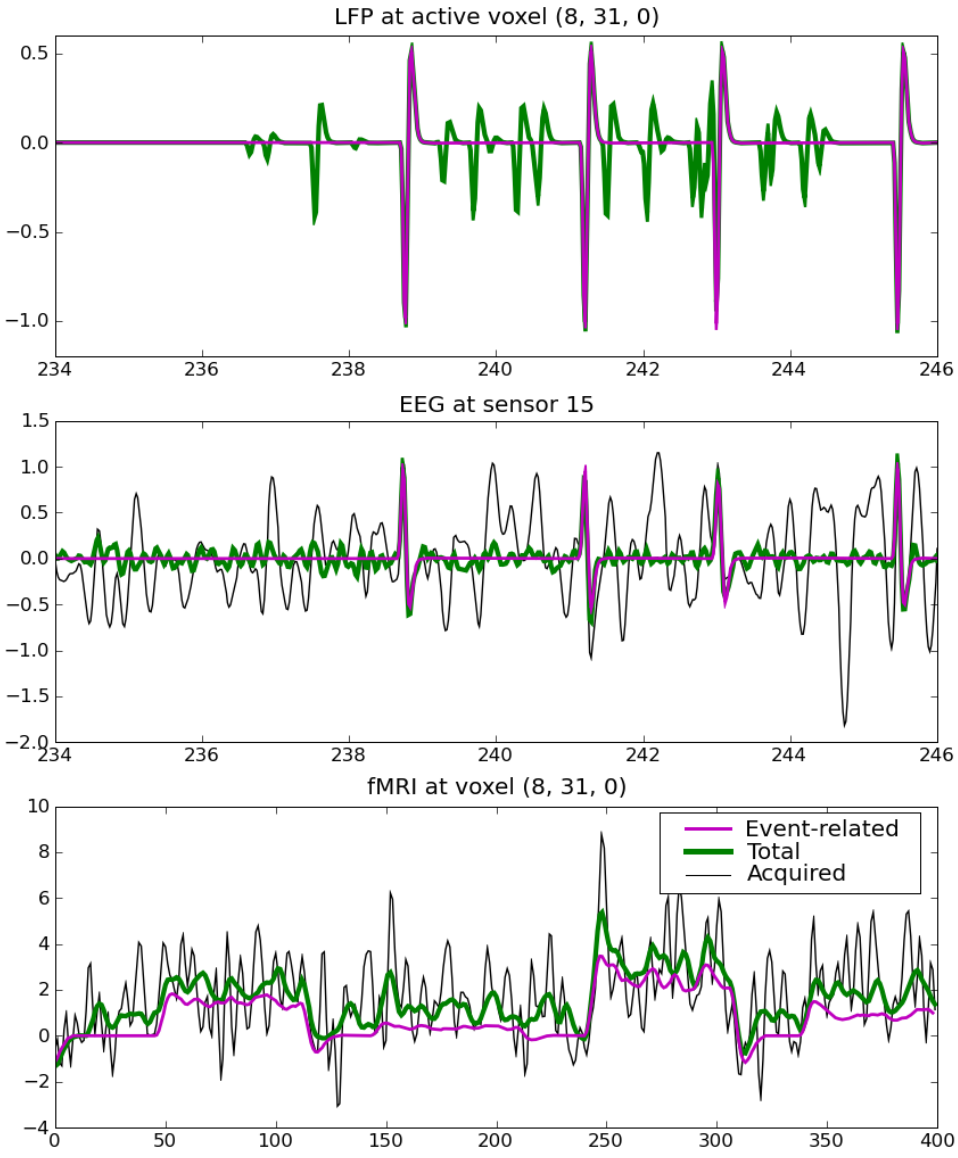


Figure 4.7 Samples of the simulated data from all data modalities at the chosen location, which carries event-related activity. EEG signal is represented by a sensor S15 which is located on the top of the "head". Event-related plots (in purple) depict simulated signal which is caused solely by ER activity. Total signal (in green) shows total signal caused by all neural activity, which includes both ER and spontaneous components. Acquired signal (in black) shows total signal with additive noise, which should mimic the real data obtained empirically.

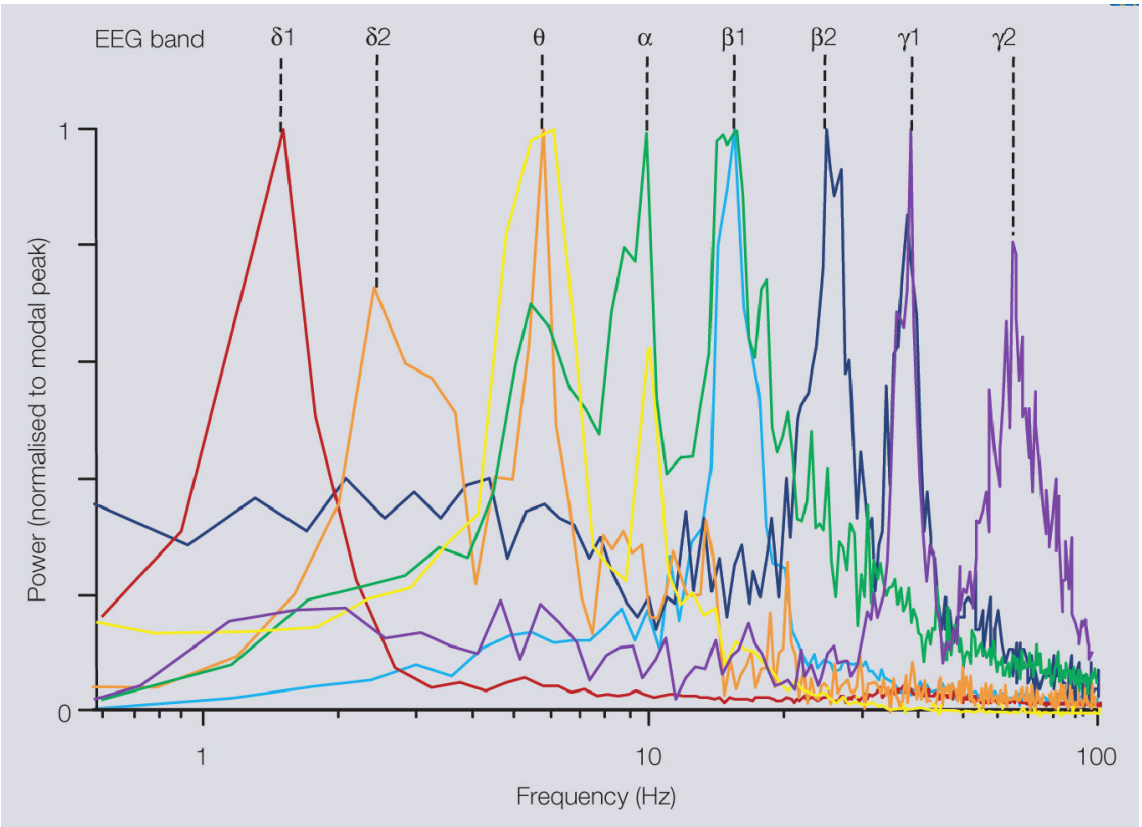


Figure 4.8 Multiple modal peak frequencies of persistent rhythms generated in isolated neocortex in vitro. (Adapted from [218, Figure 1].)

$20(\text{time-bins}) \times 9(\text{spectral-bins}) \times 16(\text{electrodes}) = 2880$ features in EEG dataset, which had the same number of samples as the number of simulated fMRI volumes.

Each feature of EEG, and each voxel in fMRI were independently standardized (z-scored) prior the analysis. A standard 4-fold cross-validation procedure was run per each simulated voxel using linear SVR (Support Vector ϵ-insensitive regression) on a noisy data (EEG and fMRI) to obtain estimates of how well this regression could predict unseen fMRI data.

4.5.3 Signal Reconstruction

Reconstruction performance of time-series was characterized by correlation coefficient between target (not observed during training) and predicted data. Goodness of prediction of noisy unseen data across all voxels of the brain is presented on Figure 4.9. Image on the left shows only voxels which passed thresholding at z=3.0 relative to by-chance performance. This should result only in 0.1% false positives in the results (*i.e.*, on average 2 voxels from the population at hands). To assess distribution of by-chance performances, thresholding relied on the assumption that the histogram of performances is a mixture of two distributions: normal distribution centered at 0 which corresponds to by-chance reconstruction performance, and the other, arbitrarily shaped, distribution of valid reconstruction performances, which does not have any significant number of negative values. Such assumption readily allows to assess the standard deviation of the distribution for by-chance performance simply by symmetrizing negative performance values around 0 and computing the standard deviation of the obtained distribution.

Since this analysis is done on the simulated data, it is more valuable to judge the performance of the method by inspecting the reconstruction relative to the noiseless data. Figure 4.10 shows that the reconstruction performance is significantly higher when

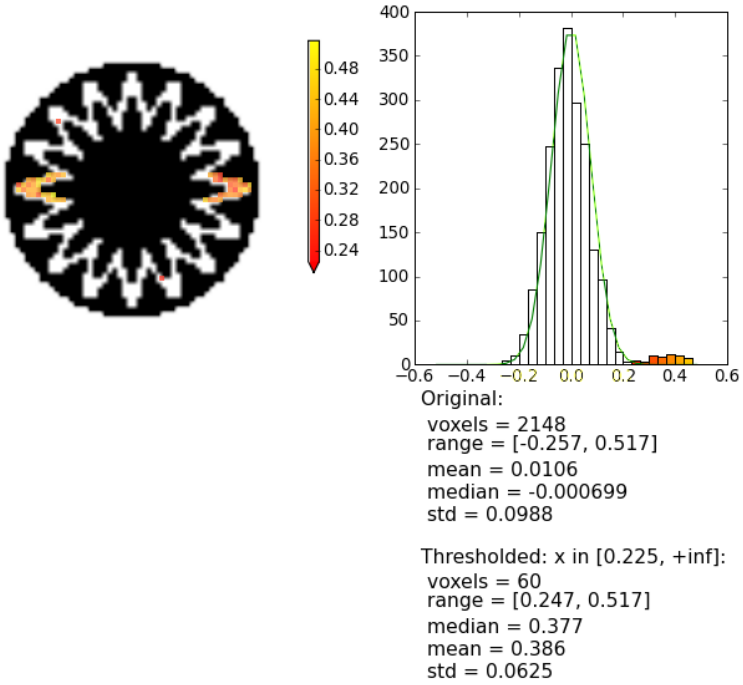


Figure 4.9 Thresholded correlation coefficients between noisy and predicted fMRI data.

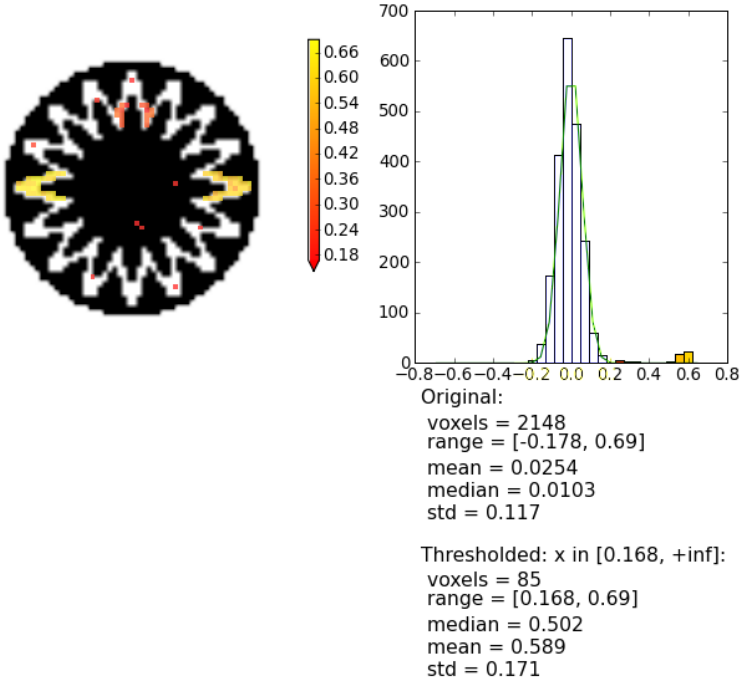


Figure 4.10 Thresholded correlation coefficients between clean and predicted fMRI data.

compared to the noiseless data, and that the distribution of by-chance performances is nicely segregated from the distribution of the meaningful reconstructions.

To visually inspect the goodness of reconstruction, time series for four chosen representative voxels are presented on Figure 4.11. These voxels correspond to a single voxel per each ROI (randomly chosen within ROIs) and a voxel at midbrain. Activity of the midbrain voxel should not possibly be reflected in EEG signal, thus neither reliably reconstructed. Figure 4.11 shows that time series within ROI1 and ROI2 were relatively well reconstructed, especially in respect to the clean original signal, although the learner was exposed during training only to the noisy EEG and fMRI data.

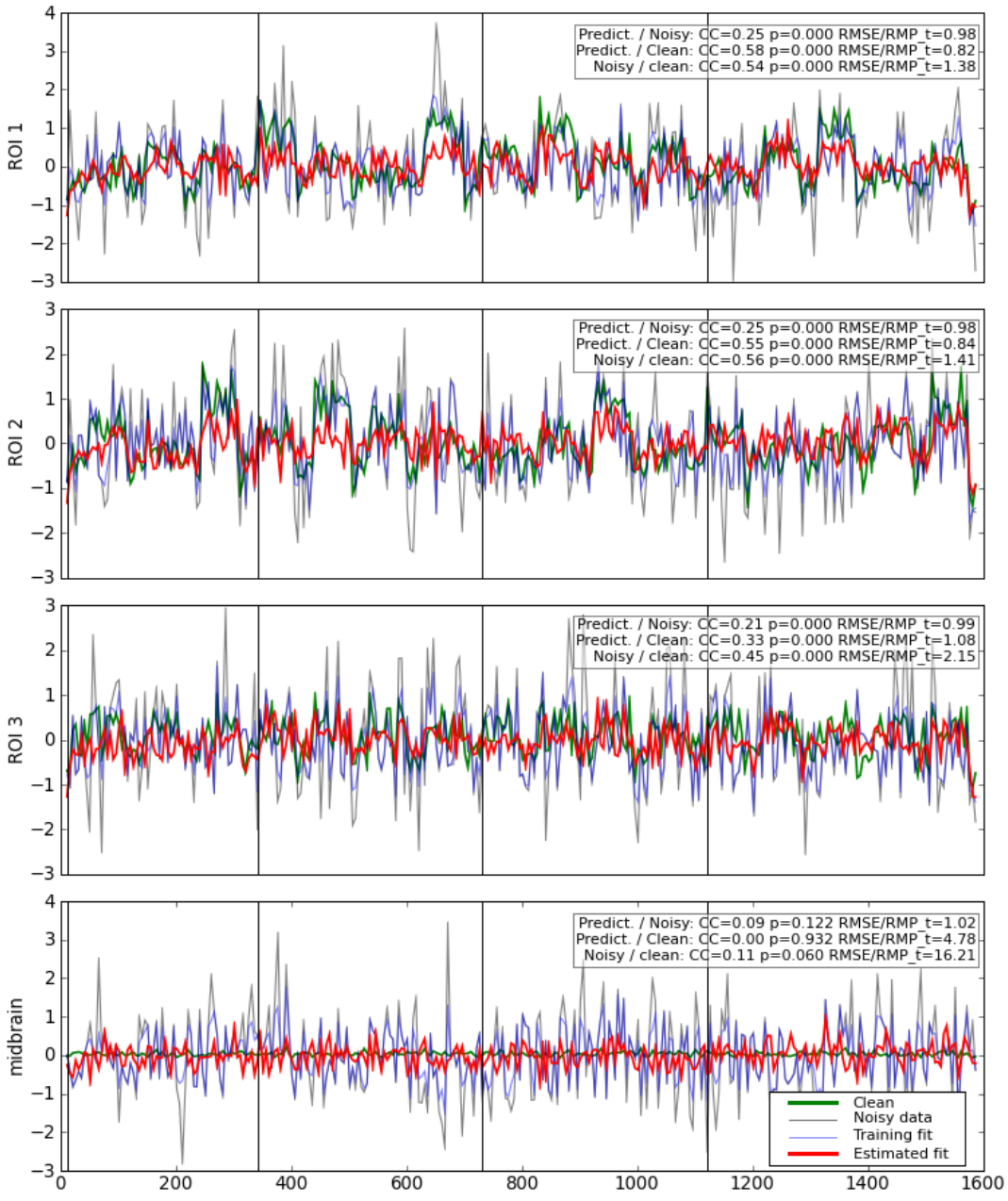


Figure 4.11 Original and predicted fMRI time series for representative voxels. CC stands for correlation coefficient (or Pearson's correlation) with a corresponding significance level p. RMSE/RMS_t is the ratio between root-mean squared error (RMSE) and root-mean square of the target signal (RMS_t). Perfect reconstruction would correspond to CC=1.0 and RMSE/RMS_t=0.0.

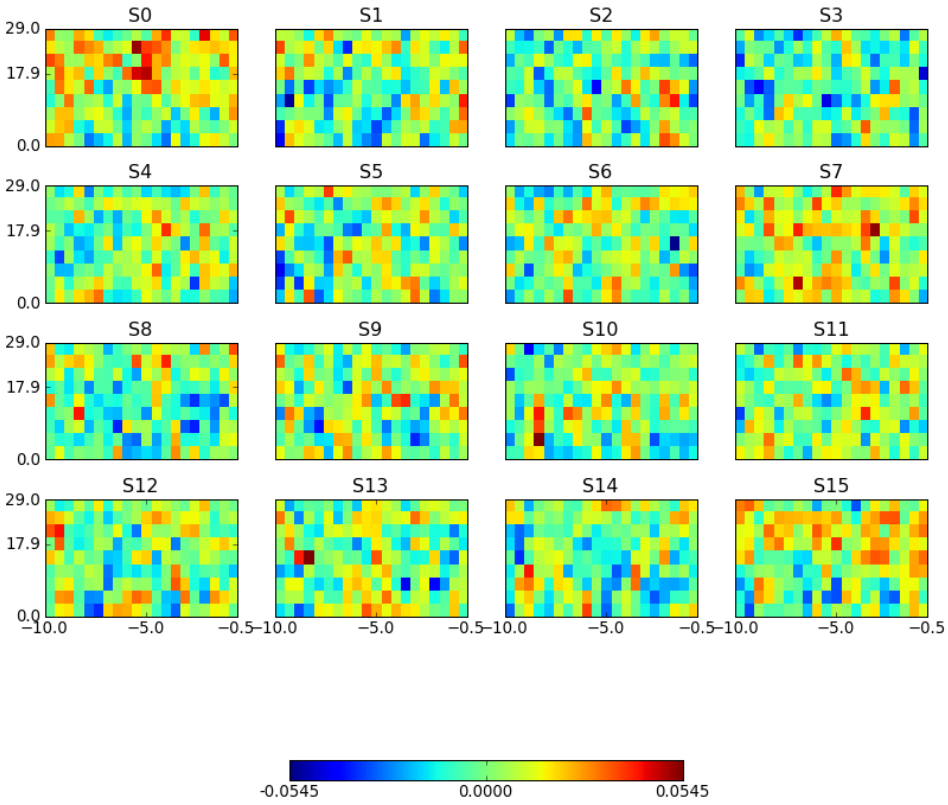


Figure 4.12 Sensitivities of SVR trained to predict fMRI signal of a voxel in ROI 1. Each plot corresponds to a single channel of EEG and represents sensitivities in time (horizontal) and frequency (vertical) bins.

4.5.4 Sensitivity Analysis

Demonstrated reconstruction of the original time-series within the ROI voxels allows to state, that the regression has learned information about neural activation which was embedded in both EEG and fMRI signals. It allows further to inspect the sensitivities of the learned regression. Figure 4.12 shows regression's sensitivity for the representative voxel within ROI1, which is neighboring S0 and S7 electrodes of EEG. Sensitivities show that results are highly sensitive to the high frequencies components of the S0.

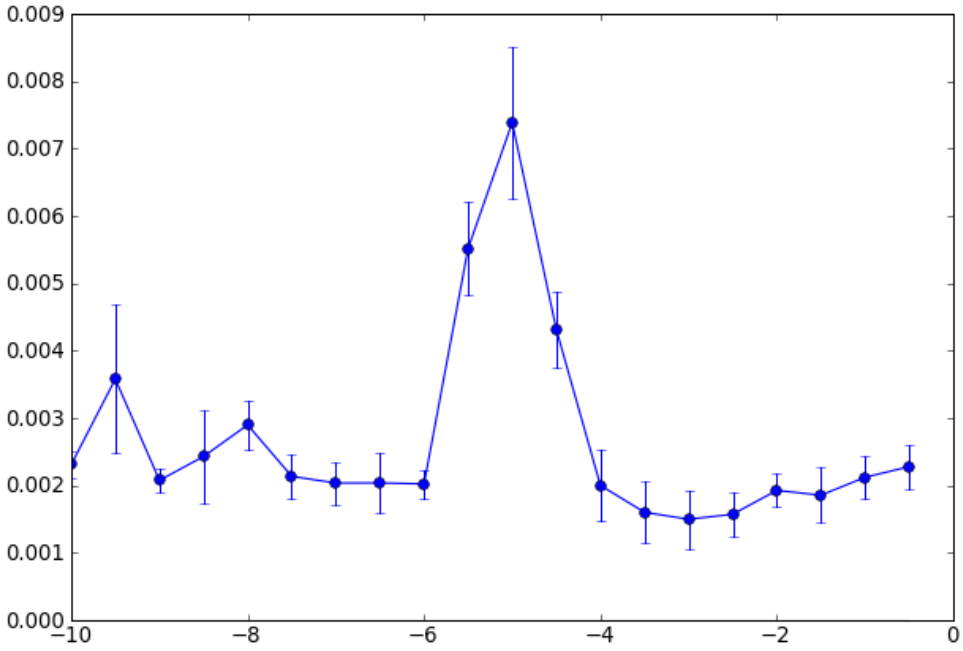


Figure 4.13 Aggregate sensitivity of SVR to the high-frequencies of S0 EEG electrode while predicting a voxel in ROI1. Error-bars correspond to the standard error estimate across 4 splits of data.

Aggregate sensitivity (sum of powers of the sensitivities) in the range of high-frequencies for that sensor is presented in Figure 4.13. Maximum sensitivity nicely matches the time-lag of 5 sec where HRF used for fMRI simulation reaches its maximum value (see Figure C.2). Absence of the sensitivity to the lower frequencies of EEG signal could be attributed to the fact that simulated instrumental noise added to EEG signal was low-pass filtered, hence contributing mostly to the lower frequencies, thus making them less reliable for the estimation of the BOLD response. Obtained results allow to conclude that the analysis of the regression sensitivities provides means to assess characteristics of the underlying HRF of the BOLD response.

4.6 Application to Real EEG/FMRI Data

The analysis of the simulated data presented in the previous section has validated the viability of the suggested method. This section presents the results of applying exactly the same analysis workflow to real data. Data from a single subject for this analysis was kindly provided by Dr.Herrmann. Details about data acquisition setup, preprocessing stages, and the results of conventional analysis using GLM and ERP methods could be found in the original publication [217].

4.6.1 Data Preparation

EEG and fMRI data were acquired simultaneously in a single session, while a subject was performing a simple auditory task at different levels of stimulation. FMRI data consisted of 147 volumes of 4 slices each (thickness 8 mm), which were acquired at the in-plane resolution of 3 mm. Due to dis-synchronization of MRI and EEG acquisition equipments, nominal fMRI TR of 10.696 sec was adjusted to be 10.7316 sec to match the time clock of EEG acquisition hardware. Such TR was deduced by the analysis EEG data, which carries heavy artifacts which occur whenever MRI acquisition is in progress (see Section 4.1.1). FMRI data was preprocessed using FSL tools: spatially smoothed at FWHM of 6 mm and temporally high-pass filtered at the cutoff corresponding to 140 sec. Only brain voxels were selected for the analysis.

EEG data was corrected to remove MR-artifacts (see [217] for the details), temporally low-pass filtered at cut-off frequency of 20 Hz and downsampled at sampling frequency of 50 Hz to match temporal resolution of the simulated data. All EEG data was re-referenced to the mean between TP9 and TP10 electrodes (see Figure C.4 for the electrodes placement in a typical EEG system). 10 out of 29 EEG sensors (Cz, CP1, CP2, Pz, T7, T8, TP9, TP10, P7, P8) were selected to reduce the input dimensionality of

the data. These sensors were selected as representative for providing clean ERP response to the stimuli.

4.6.2 Signal Reconstruction

Analysis workflow was the same as for the simulated data in Section 4.5.3 with a single difference. In this dataset TR was relatively high (10 sec), which means that actual acquisition of the slices was spread through that duration. Not to rely on the sequence of slice acquisition, $\tau = 20$ sec was taken to integrate amount of information sufficient for accounting for fMRI data in all four slices.

Reconstruction performance of the data was statistically significant across a variety of brain regions (Figure 4.14). Time series of four chosen voxels from different parts of the brain are shown in Figure 4.16. Results are surprising in the sense that the reconstruction achieved relatively high performance not only in the exterior areas, which are located in vicinity of EEG electrodes, but also in medial areas which, presumably, should be weakly represented in EEG signals registered on the scalp. This effect can be attributed to the present strong activation in those areas while performing a given task. Indeed, all the areas detected with GLM analysis (Figure 4.15) are also present on the thresholded reconstruction map (Figure 4.14). Additionally, left auditory cortex which was not significantly active according to the GLM results, was reconstructed at a comparable level with the right auditory cortex. A possible explanation could be a modulated activation in the left auditory cortex, so that GLM analysis which aims at detecting consistent activation across the trials, is not able to account for such variance, hence it is incapable to detect the activation.

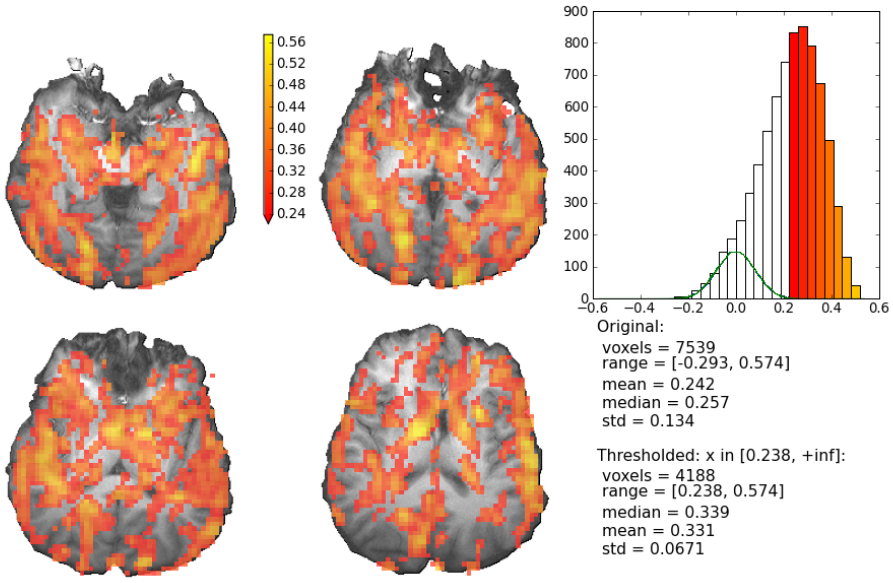


Figure 4.14 Thresholded correlation coefficients between real and predicted fMRI data in a single subject from [217].

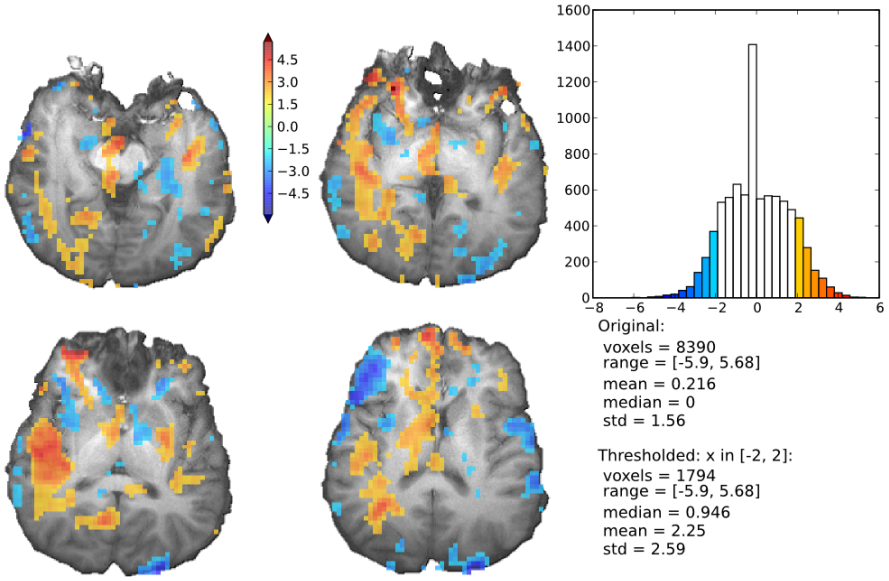


Figure 4.15 GLM analysis: significant (thresholded at $|z| \leq 2.0$) activation (in red) and deactivation (in blue) in response to auditory stimulation. Slices are plotted in radiological convention, hence left side of the brain is presented to the right from inter-hemispheric point.

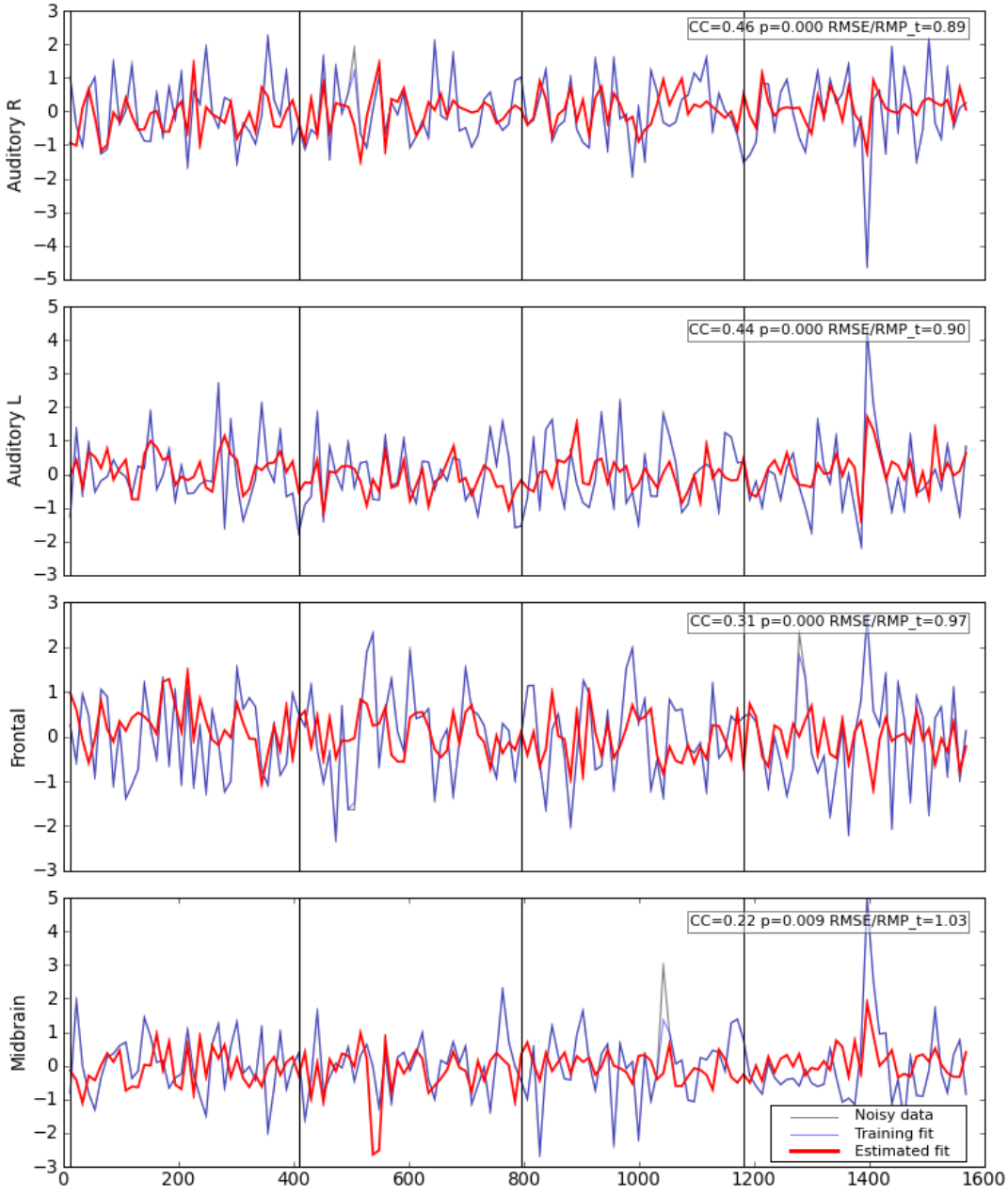


Figure 4.16 Original and predicted fMRI time series for representative voxels.

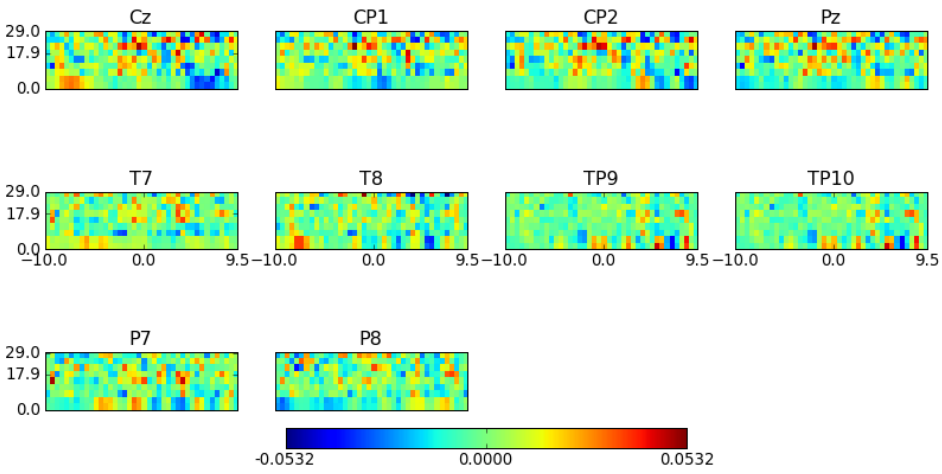


Figure 4.17 Sensitivities of SVR trained to predict fMRI signal of a voxel with right auditory cortex. Each plot corresponds to a single channel of EEG and represents sensitivities in time (horizontal) and frequency (vertical) bins.

4.6.3 Sensitivity Analysis

Figure 4.17 shows the sensitivity estimates of the SVR for a single voxel. Clearly there is higher amount of variance in sensitivities compared the sensitivities obtained on the simulated data (Figure 4.12). Nevertheless, sensitivities to the signal components at lower frequencies seems to be consistent with the assumption of the delayed BOLD response. Aggregate sensitivity for low frequencies of the sensor Cz is presented in Figure 4.18. Since for this analysis, EEG signal of duration covering acquisition of two fMRI volumes (20 sec) was used as input to the regression, it seems to be logical that fMRI signal at a given time point could be sensitive to the activity which is reflected in the previous or in the next volume (depending on the slice position).

Many studies (see Sections 4.3.1 and 4.3.2) aim to discover the brain regions which are generators of specific rhythms of cortical activity. Analysis of the SVR sensitivities to different spectral bands could allow to detect the areas of the brain where a particular frequency band contributes the most variance to the reconstruction of fMRI BOLD signal.

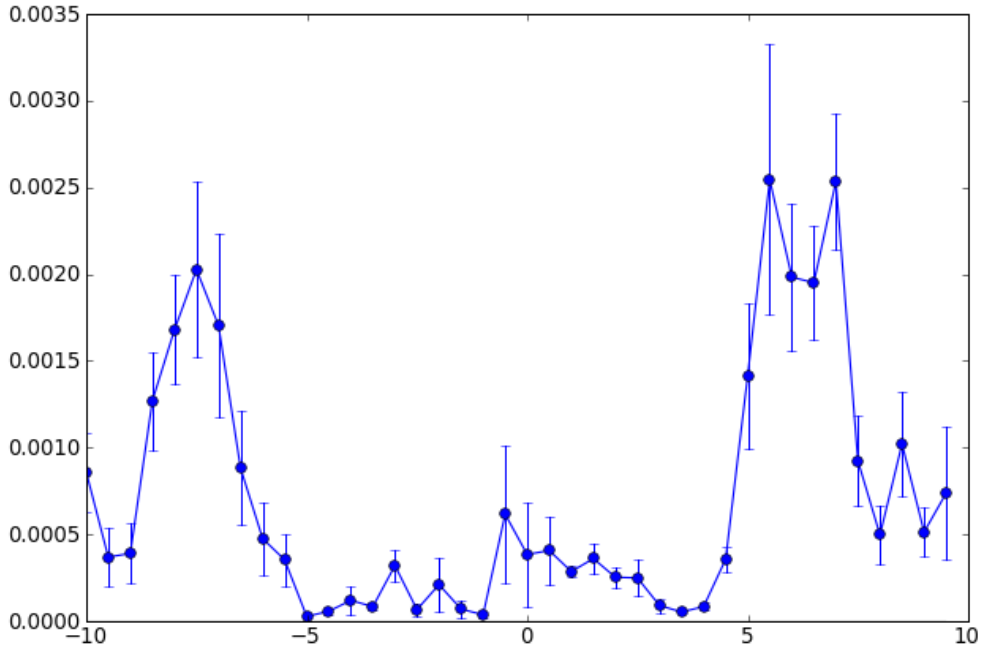


Figure 4.18 Aggregate sensitivity of SVR to the low-frequencies of Cz EEG electrode while predicting a voxel in the right auditory cortex.

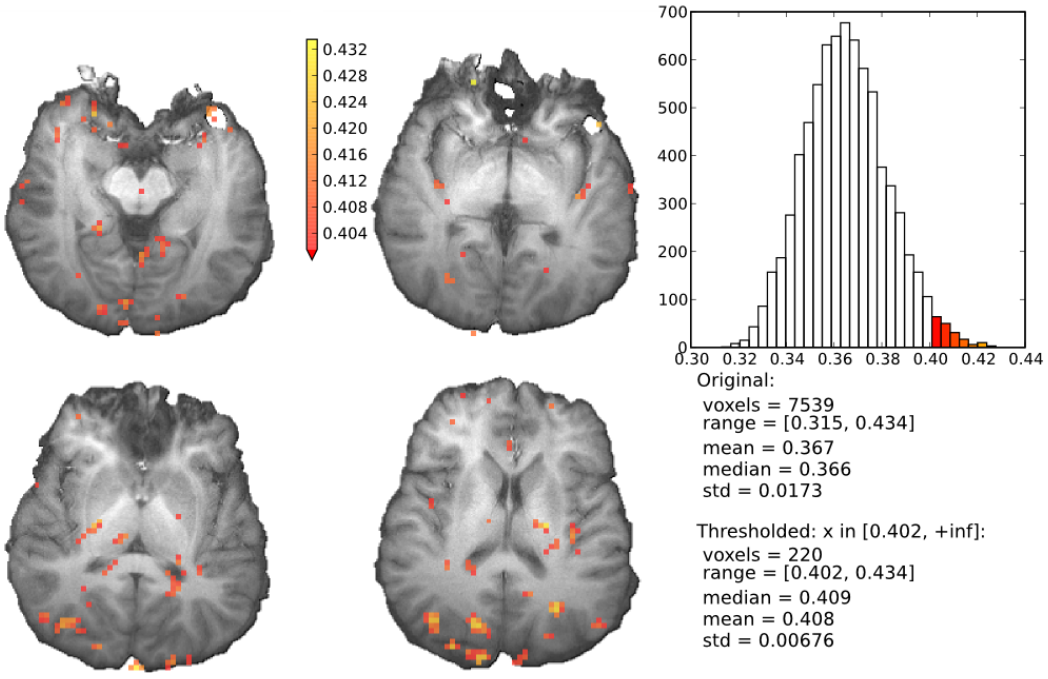


Figure 4.19 Aggregate sensitivity to the frequencies in the α-band range of frequencies.

Figure 4.19 shows thresholded ($z > 2.0$ under assumption of normally distributed values) aggregate sensitivities to the EEG signal spectral components in the α-band. Highest sensitivities are found in the occipital areas. Such finding agrees with previous studies [92, 187]. Unfortunately, limited number of the slices in the available fMRI data does not cover the full brain, therefore makes it impossible to explore all the brain regions in regards to the localization of generators of different frequency bands. Thresholded sensitivities for other frequency bands from this analysis can be found in the Appendix C (Figures C.5, C.6, and C.7).

Another possible dimension for the sensitivities aggregation is the sensitivity plots of the voxels per each EEG sensor. Figure 4.20 shows thresholded aggregated sensitivity of different voxels to the signal of T8 EEG sensor, which was positioned in the neighborhood of the temporal cortex (therefore index "T" in the name). Results show that information from T8 sensor primarily targets description of the areas such as

auditory cortex which are located in sensor's vicinity. Nevertheless, some other distant areas are also sensitive to the data in T8. For instance, parietal regions, which according to GLM results (see Figure 4.15) are not experiment-related, nevertheless, carry BOLD signal which is well described by EEG (see Figure 4.14). Whether those regions are just weakly reflected in T8 channel or actually activate coherently with auditory cortex remains an open question. Since it is unlikely for T8 electrode to pick up significant amount of variance from distant areas of the contra-lateral hemisphere, one could also speculate that supra-thresholded voxels contra-lateral to the cite of T8 are the entry points to the cortical structure from the corpus collosum[5] which facilitates the communication between two hemispheres. Hence such simple analysis of the constructed mapper sensitivities can possibly allow to deduce functional connectivity between distant areas. Aggregate sensitivities of other electrodes also seems to be in accordance with their spacial locations. For example, significant sensitivities of Pz, CP1, and CP2 electrodes cover parietal areas (Figures C.8, C.9, C.10 in Appendix C).

4.7 Conclusion

This chapter presented a novel approach toward the analysis of multiple data modalities. Suggested approach did not impose any strict forward model of fMRI response, which makes is more appropriate over existing methods. Results of the analysis of simulated and read data showed possibility to create a reliable mapping of EEG data into the space of fMRI signals. Such mapping allowed to identify the regions of the brain which were most active during the experiment but were not detected by the conventional analysis of fMRI data. Furthermore, analysis of the constructed mapping allowed to identify the regions related to the specific rhythms of EEG signal as well as to discover possible functional connections from the areas in vicinity of EEG sensors to some distant areas.

[5] http://en.wikipedia.org/wiki/Corpus_collosum

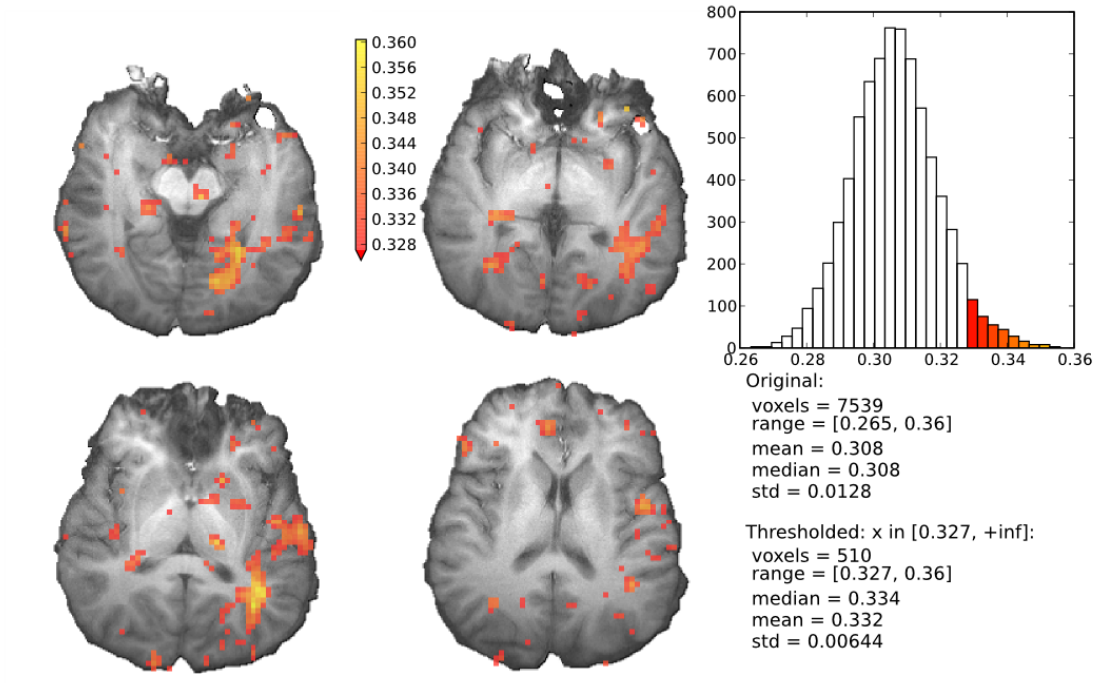


Figure 4.20 Aggregate sensitivity to the signal of T8 EEG channel (see Figure C.4 for the typical location of the sensor in respect to the brain).

CHAPTER 5

PYMVPA: MACHINE LEARNING FRAMEWORK FOR NEUROIMAGING

Recently founded *Machine learning open source software*[1] project shows an impressive, nonetheless still incomplete, sample of available software packages which implement ML methods. At the very least starting with these already available high-quality software libraries has the potential to accelerate scientific progress in the emerging field of classifier-based analysis of brain-imaging data. Although these libraries are freely available, their usage typically assumes a high-level of programming expertise and statistical or mathematical knowledge.

Therefore, it would be of great value to have a unifying framework that helps to bridge well-established neuroimaging tools and machine learning software packages and provides ease of programmability, cross library integration and transparent fMRI data handling. Such framework should at least have the five following features:

User-centered programmability with an intuitive user interface Since most neuroimaging researchers are not also computer scientists, it should require only a minimal amount of programming ability. Workflows for typical analyses should be supported by a high-level interface that is focused on the experimental design and language of the neuroimaging scientist. That being said, of course, all interfaces should allow access to detailed information about the internal processing for comprehensive extensibility. Finally, reasonable documentation is a primary requirement.

Extensibility It should be easy to add support for additional external machine learning toolboxes to prevent duplicating the effort that is necessary when a single algorithm has to be implemented multiple times.

[1] http://www.mloss.org

Transparent reading and writing of datasets Because the toolbox is focused on neuroimaging data, the default access to data, should require little or no specification for the user. The toolbox framework should also take care of proper conversions into any target data format required for the external machine learning algorithms.

Portability It should not impose restrictions about hardware platforms and should be able to run on all major operating systems.

Open source software It should be open source software, as it allows one to access and to investigate every detail of an implementation, which improves the reproducibility of experimental results, leading to more efficient debugging and gives rise to accelerated scientific progress [16].

PyMVPA, a Python-based toolbox for multivariate pattern analysis of neuroimaging data, was designed to meet all the above criteria for a classifier-based analysis framework. Following sections provide only a rough overview about the most important aspects of the toolbox. However, a comprehensive user manual is available [29] to provide more details and typical code examples.

5.1 Design

One of the main goals of PyMVPA is to reduce the gap between the neuroscience and ML communities. To reach this goal, PyMVPA was designed to provide a convenient, easy to use, community developed (free and open source[2]), and extensible framework to facilitate use of ML techniques on neural information. PyMVPA combines Python data processing, visualization, and basic I/O facilities together with I/O code and examples tailored for

[2]PyMVPA is distributed under an MIT license, which complies with both Free Software and Open Source definitions. The software is freely available in source and in binary form from the project website http://www.pymvpa.org

neuroscience. Table D.2 lists a number of Python modules which might be of interest in the neuroscientific context. For an easy start into PyMVPA an fMRI example dataset [a single subject from the study by 25] is available for download from the PyMVPA website.

As Table D.2 highlighted, PyMVPA is not the only ML framework available for scripting and interactive data exploration in Python. In contrast to some of the primarily GUI-based ML toolboxes (*e.g.*, Orange, Elephant), PyMVPA is designed to provide not just a toolbox, but a framework for concise, yet intuitive, scripting of possibly complex analysis pipelines. To achieve this goal, PyMVPA provides a number of building blocks that can be combined in a very flexible way.

Figure 2.1 in Section 2.1 showed a schematic representation of the framework design, its building blocks and how they can be combined into complete analysis pipelines. PyMVPA consists of several components (gray boxes) such as ML algorithms or dataset storage facilities. Each component contains one or more modules (white boxes) providing a certain functionality, *e.g.*, *classifiers*, but also feature-wise measures [*e.g.*, I-RELIEF; 34], and feature selection methods [recursive feature elimination, RFE; 35, 36]. Typically, all implementations within a module are accessible through a uniform interface and can therefore be used interchangeably, *i.e.*, any algorithm using a classifier can be used with any available classifier implementation, such as *support vector machine* [SVM; 17], or *sparse multinomial logistic regression* [SMLR; 37].

Some ML modules provide generic *meta* algorithms that can be combined with the *basic* implementations of ML algorithms. For example, a *Multi-Class* meta classifier provides support for multi-class problems, even if an underlying classifier is only capable to deal with binary problems.

Additionally, PyMVPA makes use of a number of external software packages (black boxes in Figure 2.1), including other Python modules and low-level libraries and computing environments (*e.g.*, *R*[3]). In the case of *SVM*, classifiers are interfaced to the

[3] http://www.r-project.org

implementations in *Shogun* or *LIBSVM*. PyMVPA only provides a convenience wrapper to expose them through a uniform interface. Using externally developed software instead of reimplementing algorithms has the advantage of a larger developer and user base and makes it more likely to find and fix bugs in a software package to ensure a high level of quality.

However, using external software also carries the risk of breaking functionality when any of the external dependencies break. To address this problem PyMVPA utilizes an automatic testing framework performing various types of tests ranging from unittests (currently covering 84% of all lines of code) to sample code snippet tests in the manual and the source code documentation itself to more evolved "real-life" examples. This facility allows one to test the framework within a variety of specific settings, such as the unique combination of program and library versions found on a particular user machine.

At the same time, the testing framework also significantly eases the inclusion of code by a novel contributor by catching errors that would potentially break the project's functionality. Being open-source does not always mean *easy to contribute* due to various factors such as a complicated application programming interface (API) coupled with undocumented source code and unpredictable outcomes from any code modifications (bug fixes, optimizations, improvements). PyMVPA welcomes contributions, and thus, addresses all the previously mentioned points:

Accessibility of source code and documentation All the source code (including website and examples) together with the full development history is publicly available via a distributed version control system, which makes it very easy to track the development of the project, as well as to develop independently and to submit back into the project.

Inplace code documentation Large parts of the source code are well documented using *reStructuredText*[4], a lightweight markup language that is highly readable in source format as well as being suitable for automatic conversion into HTML or PDF reference documentation. In fact, *Ohloh.net*[5] source code analysis judges PyMVPA as having "extremely well-commented source code".

Developer guidelines A brief summary defines a set of coding conventions to facilitate uniform code and documentation look and feel. Automatic checking of compliance to a subset of the coding standards is provided through a custom *PyLint*[6] configuration, allowing early stage minor bug catching.

Moreover, PyMVPA does not raise barriers by being limited to specific platforms. It could fully or partially be used on any platform supported by Python (depending on the availability of external dependencies). However, to improve the accessibility, convenient ways are provided for the installation across a variety of platforms: binary installers for Windows, and MacOS X, as well as binary packages for Debian GNU/Linux (included in the official repository), Ubuntu, and a large number of RPM-based GNU/Linux distributions, such as OpenSUSE, RedHat, CentOS, Mandriva, and Fedora. Additionally, the available documentation provides detailed instructions on how to build the packages from source on many platforms.

By providing simple, yet flexible interfaces, PyMVPA is specifically designed to connect to and use externally developed software. Any analysis built from those basic elements can be cross-validated by running them on multiple dataset splits that can be generated with a variety of data resampling procedures [*e.g.*, bootstrapping, 38]. Detailed information about analysis results can be queried from any building block and can be visualized with various plotting functions that are part of PyMVPA, or can be mapped

[4] http://en.wikipedia.org/wiki/ReStructuredText
[5] http://www.ohloh.net/projects/pymvpa/factoids
[6] http://www.logilab.org/projects/pylint

back into the original data space and format to be further processed by specialized tools (*i.e.*, to create an overlay volume analogous to a statistical parametric mapping). The solid arrows in Figure 2.1 represent a typical connection pattern between the modules. Dashed arrows refer to additional compatible interfaces which, although potentially useful, are not necessarily used in a standard processing chain.

A final important feature of PyMVPA is that it allows, by design, researchers to compress complex analyses into a small amount of code. This makes it possible to complement publications with the source code actually used to perform the analysis as supplementary material. Making this critical piece of information publicly available allows for in-depth reviews of the applied methods on a level well beyond what is possible with verbal descriptions. To demonstrate this feature, this paper is accompanied by the full source code to perform all analyses shown in the following sections.

5.2 Dataset Handling

Input, output, and conversion of datasets are a key task for PyMVPA. A dataset representation has to be simple enough to allow for maximum interoperability with other toolkits, but simultaneously also has to be comprehensive in order to make available as much information as possible to *e.g.*, domain-specific analysis algorithms.

One of the key features of PyMVPA is its ability to read fMRI datasets and transform them into a generic format that makes it easy for other data processing toolboxes to inherit them. PyMVPA comes with a specialized dataset type for handling import from and export to images in the NIfTI or inferior ANALYZE formats. It automatically configures an appropriate mapper by reading all necessary information from the NIfTI file header. Upon export, all header information is preserved (including

embedded transformation matrices). This makes it very easy to do further processing in any other NIfTI-aware software package, like *AFNI*[7], *BrainVoyager*[8], *FSL*[9] or *SPM*[10].

Since many algorithms are applied only to a subset of voxels, PyMVPA provides convenient methods to select voxels based on ROI masks. Successively applied feature selections will be taken into account by the mapping algorithm of NIfTI datasets and reverse mappings from the new subspace of features into the original dataspace, *e.g.*, for visualization, is automatically performed upon request.

However, the mapper construct in PyMVPA, which is applied to each data sample, is more flexible than a simple 3-D data volume to 1-D feature vector transformation. The original dataspace is not limited to three dimensions. For example, when analyzing an experiment using an event-related paradigm it might be difficult to select a single volume that is representative for some event. A possible solution is to select all volumes covering an event in time, which results in a four-dimensional dataspace. A mapper can also be easily used for EEG/MEG data, *e.g.*, mapping spectral decompositions of the time series from multiple electrodes into a single feature vector. PyMVPA provides convenient methods for these use-cases and also supports reverse mapping of results into the original dataspace, which can be of any dimensionality.

5.3 Machine Learning Algorithms

PyMVPA itself does not at present implement all possible classifiers, even if that were desirable. Currently included are implementations of a k-nearest-neighbor classifier as well as ridge, penalized logistic, Bayesian linear, Gaussian process (GPR), and sparse multinomial logistic regressions [SMLR; 37]. However, instead of distributing yet another implementation of popular classification algorithms the toolbox defines a generic classifier

[7] http://afni.nimh.nih.gov/afni
[8] http://www.brainvoyager.com
[9] http://www.fmrib.ox.ac.uk/fsl
[10] http://www.fil.ion.ucl.ac.uk/spm

interface that makes it possible to easily create software wrappers for existing machine learning libraries and enable their classifiers to be used within the PyMVPA framework. At the time of this writing, wrappers for support vector machine algorithms [SVM; 17] of the widely used *LIBSVM* package [219] and *Shogun* machine learning toolbox [220] are included. Additional classifiers implemented in the statistical programming language R are provided within PyMVPA [*e.g.*, least angle regression, LARS, 221]. The software wrappers expose as much functionality of the underlying implementation as necessary to allow for a seamless integration of the classification algorithm into PyMVPA. Wrapped classifiers can be treated and behave exactly as any of the native implementations.

Some classifiers have specific requirements about the datasets they can be trained on. For example, support vector machines [17] do not provide native support for multi-class problems, *i.e.*, discrimination of more than two classes. To deal with this fact, PyMVPA provides a framework to create meta-classifiers (see Figure 2.1). These are classifiers that utilize several basic classifiers, both those implemented in PyMVPA and those from external resources, that are each trained separately and their respective predictions are used to form a joint meta-prediction, sometimes referred to as *boosting* [see 222]. Besides generic multi-class support, PyMVPA provides a number of additional meta-classifiers *e.g.*, a classifier that automatically applies a customizable feature selection procedure prior to training and prediction. Another example is a meta-classifier that applies an arbitrary mapping algorithm to the data to implement dimensionality transformation and/or reduction step, such as, principal component analysis (PCA), independent component analysis (ICA), both using implementations from *MDP*[11] or wavelet decomposition via *pywavelets*[12].

Despite its high-level interface PyMVPA offers detailed information to the user. To achieve a useful level of transparency, all classifiers can easily store any amount of

[11] http://mdp-toolkit.sourceforge.net
[12] http://www.pybytes.com/pywavelets/

additional information. For example, a logistic regression might optionally store the output values of the regression that are used to make a prediction. PyMVPA provides a framework to store and pass this information to the user if it is requested. The type and size of such information is in no way limited. However, if the use of additional computational or storage resources is not required, then it can be switched off at any time, to allow for an optimal tradeoff between transparency and performance.

This section presents only a brief overview of PyMVPA organization and coding approaches. Significantly broader and in-depth coverage of PyMVPA is available from the "PyMVPA User Manual" [29]. The main goal of the section is to demonstrate the flexibility of PyMVPA for the construction of various analysis pipelines; and also to highlight the conciseness of the actual code needed to perform a desired statistical learning analysis. That theoretically should allow shipment of complete analysis code with any publication as supplemental material (*e.g.*, see [23]).

5.4 Example Analyses

This section will demonstrate the functionality of PyMVPA by running some selected analyses on fMRI data from a single participant (participant 1) from a dataset published by Haxby et al. [25]. This dataset was chosen for demonstration examples because, since its first publication, it has been repeatedly reanalyzed [13, 26, 28], was analyzed on the course of this dissertation work (see Section 3.4.2 for the description of the experiment and data preprocessing), and is readily available from PyMVPA's website.

For the sake of simplicity, analysis was focused on the binary CATS vs. SCISSORS classification problem. These two categories were chosen for the demonstration of PyMVPA facilities since their classification in general poses harder problem, than of more "specialized" ones, such as FACE and HOUSE (analysis on which was presented in the Section 3.4.3). All volumes recorded during either CATS or

SCISSORS blocks were extracted and voxel-wise Z-scored with respect to the mean and standard deviation of volumes recorded during rest periods. Z-scoring was performed individually for each run to prevent any kind of information transfer across runs.

Every analysis is accompanied by source code snippets that show their implementation using the PyMVPA toolbox. For demonstration purposes they are limited to the most important steps.

5.4.1 Loading a Dataset

Dataset representation in PyMVPA builds on NumPy arrays. Anything that can be converted into such an array can also be used as a dataset source for PyMVPA. Possible formats range from various plain text formats to binary files. However, the most important input format from the functional imaging perspective is NIfTI[13], which PyMVPA supports with a specialized module.

The following short source code snippet demonstrates how a dataset can be loaded from a NIfTI image. PyMVPA supports reading the sample attributes from a simple two-column text file that contains a line with a label and a chunk id for each volume in the NIfTI image (line 0). To load the data samples from a NIfTI file it is sufficient to create a `NiftiDataset` object with the filename as an argument (line 1). The previously-loaded sample attributes are passed to their respective arguments as well (lines 2-3). Optionally, a mask image can be specified (line 4) to easily select a subset of voxels from each volume based on the non-zero elements of the mask volume. This would typically be a mask image indicating brain and non-brain voxels.

```
0 attr = SampleAttributes('sample_attr_filename.txt')
1 dataset = NiftiDataset(samples='subj1_bold.nii.gz',
2                        labels=attr.labels,
3                        chunks=attr.chunks,
4                        mask='subj1_roi_mask.nii.gz')
```

[13] To a certain degree PyMVPA also supports importing ANALYZE files.

Once the dataset is loaded, successive analysis steps, such as feature selection and classification, only involve passing the `dataset` object to different processing objects. All following examples assume that a dataset was already loaded.

5.4.2 Simple Full-brain Analysis

The first analysis example shows the few steps necessary to run a simple cross-validated classification analysis. After a dataset is loaded, it is sufficient to decide which classifier and type of splitting shall be used for the cross-validation procedure. Everything else is automatically handled by `CrossValidatedTransferError`. The following code snippet performs the desired classification analysis via leave-one-out cross-validation. Error calculation during cross-validation is conveniently performed by `TransferError`, which is configured to use a linear C-SVM classifier[14] on line 6. The leave-one-out cross-validation type is specified on line 7.

```
5 cv = CrossValidatedTransferError(
6         transfer_error=TransferError(LinearCSVMC()),
7         splitter=NFoldSplitter(cvtype=1))
8 mean_error = cv(dataset)
```

Simply passing the dataset to `cv` (line 8) yields the mean error. The computed error defaults to the fraction of incorrect classifications, but an alternative error function can be passed as an argument to the `TransferError` call. If desired, more detailed information is available, such as a confusion matrix based on all the classifier predictions during cross-validation.

[14] LIBSVM C-SVC [219] with trade-off parameter C being a reciprocal of the squared mean of Frobenius norms of the data samples.

5.4.3 Multivariate Searchlight

One method to localize functional information in the brain is to perform a classification analysis in a certain ROI [*e.g.*, 223]. The rationale for size, shape and location of a ROI can be *e.g.*, anatomical landmarks or functional properties determined by a GLM-contrast.

Alternatively, Kriegeskorte et al. [20] proposed an algorithm that scans the whole brain by running multiple ROI analyses. The so-called *multivariate searchlight* runs a classifier-based analysis on spherical ROIs of a given radius centered around any voxel covering brain matter. Running a searchlight analysis computing *e.g.*, generalization performances yields a map showing where in the brain a relevant signal can be identified while still harnessing the power of multivariate techniques [for application examples see 19, 224].

A searchlight performs well if the target signal is available within a relatively small area. By increasing size of the searchlight the information localization becomes less specific because, due to the anatomical structure of the brain, each spherical ROI will contain a growing mixture of gray-matter, white-matter, and non-brain voxels. Additionally, a searchlight operating on volumetric data will integrate information across brain-areas that are not directly connected to each other *i.e.*, located on opposite borders of a sulcus. This problem can be addressed by running a searchlight on data that has been transformed into a surface representation. PyMVPA supports analyses with spatial searchlights (not extending in time), operating on both volumetric and surface data (given an appropriate mapping algorithm and using circular patches instead of spheres). The searchlight implementation can compute an arbitrary measure within each sphere.

In the following example, the measure to be computed by the searchlight is configured first. Similar to the previous example it is a cross-validated transfer or generalization error, but this time it will be computed on an odd-even split of the dataset and with a linear C-SVM classifier (lines 9-11). On line 12 the searchlight is setup to compute this measure in all possible 5 mm-radius spheres when called with a dataset

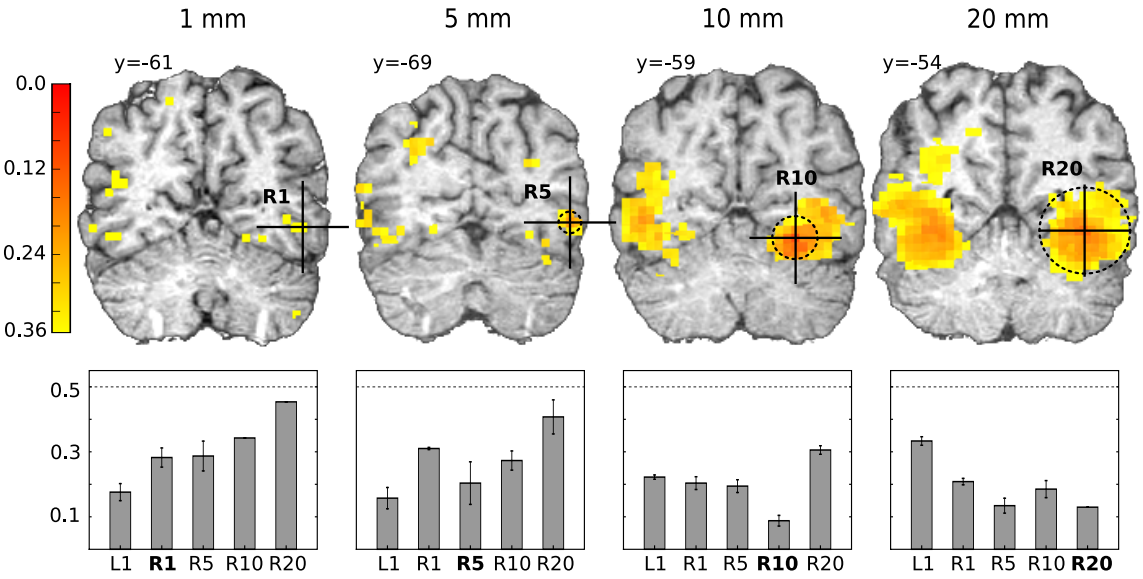


Figure 5.1 Searchlight analysis results for CATS vs. SCISSORS classification for sphere radii of 1, 5, 10 and 20 mm (corresponding to approximately 1, 15, 80 and 675 voxels per sphere respectively). The upper part shows generalization error maps for each radius. All error maps are thresholded arbitrarily at 0.35 (chance level: 0.5) and are not smoothed to reflect the true functional resolution. The center of the best performing sphere (*i.e.*, lowest generalization error) in right temporal fusiform cortex or right lateral occipital cortex is marked by the cross-hair on each coronal slice. The dashed circle around the center shows the size of the respective sphere (for radius 1 mm the sphere only contains a single voxel). MNI-space coordinates (x, y, z) in mm for the four sphere centers are: 1 mm (*R1*): (48, -61, -6), 5 mm (*R5*): (48, -69, -4), 10 mm (*R10*): (28, -59, -12) and 20 mm (*R20*): (40, -54, -8). The lower part shows the generalization errors for spheres centered around these four coordinates, plus the location of the univariately best performing voxel (*L1*: -35, -43, -23; left occipito-temporal fusiform cortex) for all radii. The error bars show the standard error of the mean across cross-validation folds.

(line 13) The final call on line 14 transforms the computed error map back into the original data space and stores it as a compressed NIfTI file. Such a file can then be viewed and further processed with any NIfTI-aware toolkit.

```
9  cv = CrossValidatedTransferError(
10         transfer_error=TransferError(LinearCSVMC()),
11         splitter=OddEvenSplitter())
12 sl = Searchlight(cv, radius=5)
13 sl_map = sl(dataset)
14 dataset.map2Nifti(sl_map).save('searchlight_5mm.nii.gz')
```

Figure 5.1 shows the searchlight error maps for the CATS vs. SCISSORS classification on single volumes from the example dataset for radii of 1, 5, 10 and 20 mm respectively. Utilizing only a single voxel in each sphere (1 mm radius), yields a generalization error as low as 17% in the best performing sphere, which is located in the left occipito-temporal fusiform cortex. With increases in the radius there is a tendency for further error reduction, indicating that the classifier performance benefits from integrating signal from multiple voxels. However, better classification accuracy is achieved at the cost of reduced spatial precision of signal localization. The distance between the centers of the best-performing spheres for 5 and 20 mm searchlights totals almost 18 mm. The lowest overall error in the right occipito-temporal cortex with 8% is achieved by a searchlight with a radius of 10 mm. The best performing sphere with 20 mm radius (12% generalization error) is centered between right inferior temporal and fusiform gyrus. It comprises approximately 700 voxels and extends from right lingual gyrus to the right inferior temporal gyrus, also including parts of the cerebellum and left lateral ventricle. It, therefore, includes a significant proportion of voxels sampling cerebrospinal fluid or white matter, indicating that a sphere of this size is not optimal given the structural organization of the brain surface. Kriegeskorte et al. [20] suggest that a sphere radius of 4 mm yields near-optimal performance. However, while this assumption might be valid for representations of object properties or low-level visual features, a searchlight of this size could miss signals related to high-level cognitive processes that involve several spatially distinct functional subsystems of the brain.

5.4.4 Feature Selection

Feature selection is a common preprocessing step that is also routinely performed as part of a conventional fMRI data analysis, *i.e.*, the initial removal of non-brain voxels. This basic functionality is provided by `NiftiDataset` as it was shown on line 4 to provide

an initial operable feature set. Likewise, a searchlight analysis also involves multiple feature selection steps (*i.e.*, ROI analyses), based on the spatial configuration of features. Nevertheless, PyMVPA provides additional means to perform feature selection, which are not specific to the fMRI domain, in a transparent and unified way.

Machine learning algorithms often benefit from the removal of noisy and irrelevant features [see 35, Section V.1. "The features selected matter more than the classifier used"]. Retaining only features relevant for classification improves learning and generalization of the classifier by reducing the possibility of overfitting the data. Therefore, providing a simple interface to feature selection is critical to gain superior generalization performance and get better insights about the relevance of a subset of features with respect to a given contrast. Table 5.1 shows the prediction error of a variety of classifiers on the full example dataset with and without any prior feature selection. For classifiers with feature selection the classifier algorithm is followed by the sensitivity measure the feature selection was based on (*e.g.*, LinSVM on 50(SVM) reads: linear SVM classifier using 50 features selected by their magnitude of weight from a trained linear SVM). Most of the classifiers perform near chance performance without prior feature selection[15], and even simple feature selection (*e.g.*, some percentage of the population with highest scores on some measure) boosts generalization performance significantly of all classifiers, including the non-linear algorithms radial basis function SVM and, kNN.

PyMVPA provides an easy way to perform feature selections. The `FeatureSelectionClassifier` is a meta-classifier that enhances any classifier with an arbitrary initial feature selection step. This approach is very flexible as the resulting classifier can be used as any other classifier, *e.g.*, for unbiased generalization

[15]Chance performance without feature selection was not the norm for all category pairs in the dataset. For example, the SVM classifier generalized well for other pairs of categories (*e.g.*, FACE vs HOUSE) without prior feature selection. Consequently, SCISSORS vs CATS was chosen to provide a more difficult analysis case.

testing using `CrossValidatedTransferError`. For instance, the following example shows a classifier that operates only on 5% of the voxels that have the highest ANOVA score across the data categories in a particular dataset. It is noteworthy that the source code looks almost identical to the example given on lines 5-8, with just the feature selection method added to it. No changes are necessary for the actual cross-validation procedure.

```
15 clf = FeatureSelectionClassifier(
16   LinearCSVMC(),
17   feature_selection=SensitivityBasedFeatureSelection(
18     sensitivity_analyzer=OneWayAnova(),
19     feature_selector=
20       FractionTailSelector(0.05, mode='select', tail='upper'))
21 cv = CrossValidatedTransferError(
22   transfer_error=TransferError(clf),
23   splitter=NFoldSplitter(cvtype=1))
24 mean_error = cv(dataset)
```

It is important to emphasize that feature selection (lines 17-19) in this case is *not* performed first on the full dataset, which could bias generalization estimation. Instead, feature selection is being performed as a part of classifier training, thus, only the actual training dataset is visible to the feature selection. Due to the unified interface, it is possible to create a more sophisticated example, where feature selection is performed via *recursive feature elimination* [35, 36]:

```
25 rfesvm = LinearCSVMC()
26 clf = SplitClassifier(
27   FeatureSelectionClassifier(
28     rfesvm,
29     feature_selection=RFE(
30       sensitivity_analyzer=
31         LinearSVMWeights(rfesvm,
32                          transformer=Absolute),
33       transfer_error=TransferError(rfesvm),
34       stopping_criterion=FixedErrorThresholdStopCrit(0.05),
35       feature_selector=
36         FractionTailSelector(0.2, mode='discard',
37                              tail='lower'),
38       update_sensitivity=True)),
39   splitter=NFoldSplitter())
```

On line 25 the main classifier is defined, which is reused in many aspects of the processing: line 28 specifies that classifier to be used to make the final prediction operating only on the selected features, line 31 instructs the sensitivity analyzer to use it to provide sensitivity estimates of the features at each step of recursive feature elimination, and on line 33 we specify that the error used to select the best feature set is a generalization error of that same classifier. Utilization of the same classifier for both the sensitivity analysis and for the transfer error computation prevents us from re-training a classifier twice for the same dataset.

The fact that the RFE approach is classifier-driven requires us to provide the classifier with two datasets: one to train a classifier and assess its features sensitivities and the other one to determine stopping point of feature elimination based on the transfer error. Therefore, the `FeatureSelectionClassifier` (line 27) is wrapped within a `SplitClassifier` (line 26), which in turn uses `NFoldSplitter` (line 39) to generate a set of data splits on which to train and test each independent classifier. Within each data split, the classifier selects its features independently using RFE by computing a generalization error estimate (line 33) on the internal validation dataset generated by the splitter. Finally, the `SplitClassifier` uses a customizable voting strategy (by default *MaximalVote*) to derive the joint classification decision.

As before, the resultant classifier can now simply be used within `CrossValidatedTransferError` to obtain an unbiased generalization estimate of the trained classifier. The step of validation onto independent validation dataset is often overlooked by the researchers performing RFE [35]. That leads to biased generalization estimates, since otherwise internal feature selection of the classifier is driven by the full dataset.

Fortunately, some machine learning algorithms provide internal theoretical upper bound on the generalization performance, thus they could be used as a `transfer_error` criterion (line 33) with RFE, which eliminates the necessity of

additional splitting of the dataset. Some other classifiers perform feature selection internally (*e.g.*, SMLR, also see figure 5.2), which removes the burden of external explicit feature selection and additional data splitting.

5.5 Conclusion

Numerous studies have illustrated the power of classifier-based analyses, harnessed with statistical learning theory, to extract information about the functional properties of the brain previously thought to be below the signal-to-noise ratio [for reviews see 21, 225].

Although the studies cover a broad range of topics from human memory to visual perception, it is important to note that they were performed by relatively few research groups. This may be due to the fact that very few software packages that specifically address classifier-based analyses of fMRI data are available to a broad audience. Such packages require a significant amount of software development, starting from basic problems, such as how to import and process fMRI datasets, to more complex problems, such as the implementation of classifier algorithms. This results in an initial overhead requiring significant resources before actual neuroimaging datasets can be analyzed.

The PyMVPA toolbox aims to be a solid base for conducting classifier-based analyses. In contrast to the *3dsvm* plugin for AFNI [31], it follows a more general approach by providing a collection of common algorithms and processing steps that can be combined with great flexibility. Consequently, the initial overhead to start an analysis once the fMRI dataset is acquired is significantly reduced because the toolbox also provides all necessary import, export and preprocessing functionality.

PyMVPA is specifically tuned towards the analysis of neural data, but the generic design of the framework allows to work in other domains equally well. The flexible dataset handling allows one to easily extend it to other data formats, while at the same time extending the mapping algorithms to represent other data spaces and metrics.

However, the key feature of PyMVPA is that it provides a uniform interface to bridge from standard neuroimaging tools to machine learning software packages. This interface makes it easy to extend the toolbox to work with a broad range of *existing* software packages, which should significantly reduce the need to recode available algorithms for the context of brain-imaging research. Moreover, all external and internal classifiers can be freely combined with the classifier-independent algorithms for *e.g.*, feature selection, making this toolbox an ideal environment to compare different classification algorithms.

The introduction of this chapter has listed portability as one of the goals for an optimal analysis framework. PyMVPA code is tested to be portable across multiple platforms, and its limiting set of essential external dependencies in turn has proven to be portable. In fact, PyMVPA only depends on a moderately recent version of Python and the NumPy package. Although PyMVPA can make use of other external software, the functionality provided by them is completely optional.

Although PyMVPA aims to be especially user-friendly it does not provide a graphical user interface (GUI). The reason not to include such an interface is that the toolbox explicitly aims to encourage novel combinations of algorithms and the development of new analysis strategies that are not easily foreseeable by a GUI designer[16]. The toolbox is nevertheless user-friendly, enabling researchers to conduct highly complex analyses with just a few lines of easily readable code. It achieves this by taking away the burden of dealing with low-level libraries and providing a great variety of algorithms in a concise framework. The required skills of a potential PyMVPA user are not much different from neuroscientists using the basic building blocks needed for one of the established fMRI analysis toolkits (*e.g.*, shell scripting for AFNI and FSL command line tools, or Matlab-scripting of SPM functions).

[16]Nothing prevents a software developer from adding a GUI to the toolbox using one of the many GUI toolkits that interface with Python code, such as PyQT (http://www.riverbankcomputing.co.uk/software/pyqt) or wxPython (http://www.wxpython.org/).

Recent releases of PyMVPA added support for visualization of analysis results, such as, classifier confusions, distance matrices, topography plots, and time-locked signals plots. However, while PyMVPA does not provide extensive plotting support it nevertheless makes it easy to use existing tools for MRI-specific data visualization. Similar to the data import PyMVPA's mappers make it also trivial to export data into the original data space and format, *e.g.*, using a reverse-mapped sensitivity volume as a statistical overlay, in the same fashion to SPM in conventional analysis. This way PyMVPA can fully benefit from the functionality provided by the numerous available MRI toolkits.

The features of PyMVPA outlined here cover only a fraction of the currently implemented functionality. More information is, however, available on the PyMVPA project website, which contains user manual [29] with an introduction into the main concepts and the design of the framework, a wide range of examples, a comprehensive module reference as a user-oriented summary of the available functionality, and finally a more technical reference for extending the framework.

The emerging field of classifier-based analysis of fMRI data is beginning to complement the established analysis techniques and has great potential for novel insights into the functional architecture of the brain. However, there are a lot of open questions how the wealth of algorithms developed by those motivated by statistical learning theory can be optimally applied to brain-imaging data (*e.g.*, fMRI-aware feature selection algorithms or statistical inference of classifier performances). The lack of a *gold standard* for classifier-based analysis demands software that allows one to apply a broad range of available techniques and test an even broader range of hypotheses. PyMVPA tries to reach this goal by providing a unifying framework that allows to easily combine a large number of basic building blocks in a flexible manner to help neuroscientists to do rapid initial data exploration and, consecutive custom data analysis. PyMVPA even facilitates the integration of additional algorithms in its framework that are not yet discovered by

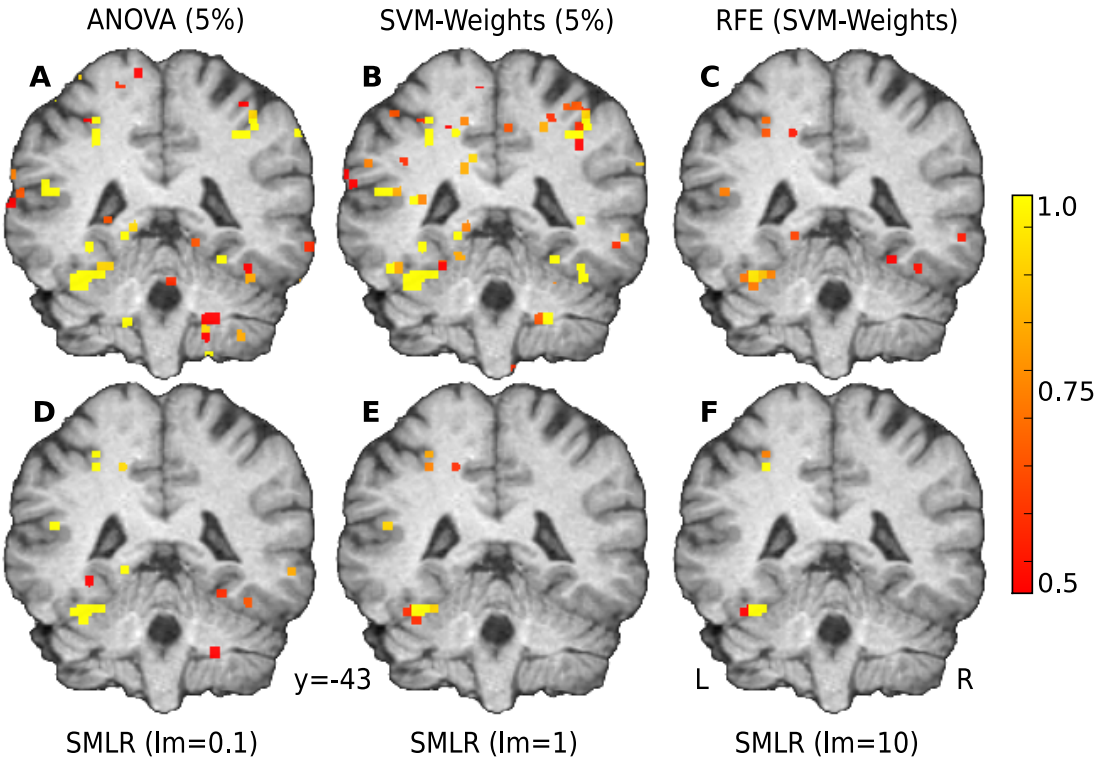


Figure 5.2 Feature selection stability maps for the CATS vs. SCISSORS classification. The color maps show the fraction of cross-validation folds in which each particular voxel is selected by various feature selection methods used by the classifiers listed in Table 5.1: 5% highest ANOVA F-scores (A), 5% highest weights from trained SVM (B), RFE using SVM weights as selection criterion (C), internal feature selection performed by the SMLR classifier with penalty term 0.1 (D), 1 (E) and 10 (F). All methods reliably identify a cluster of voxels in the left fusiform cortex, centered around MNI: -34, -43, -20 mm. All stability maps are thresholded arbitrarily at 0.5 (6 out of 12 cross-validation folds).

neuro-imaging researchers. Despite being able to perform complex analyses, PyMVPA provides a straightforward programming interface based on an intuitive scripting language. The availability of more user-friendly tools, like PyMVPA, will hopefully attract more researchers to conduct classifier-based analyses and, thus, explore the full potential of statistical learning theory based techniques for brain-imaging research.

Table 5.1 Performance of Various Classifiers with and without Feature Selection.

Classifier	Training			Transfer	
	Features utilized	Error	Time (sec)	Error (±stderr)	Time (sec)
Without feature selection					
LinSVM(C=def)	29125	0.00	22.7	0.40±0.07	2.0
LinSVM(C=10*def)	29125	0.00	22.7	0.36±0.07	2.0
LinSVM(C=1)	29125	0.00	22.6	0.36±0.07	2.0
RbfSVM()	29125	0.00	23.8	0.50±0.07	2.1
kNN()	29125	0.02	0.0	0.44±0.03	2.9
With feature selection					
SMLR(lm=0.1)	314	0.00	19.0	0.09±0.03	0.1
SMLR(lm=1.0)	92	0.00	5.0	0.11±0.03	0.1
SMLR(lm=10.0)	42	0.00	2.4	0.09±0.03	0.1
RbfSVM on SMLR(lm=10) non-0	42	0.00	2.5	0.11±0.02	0.0
kNN on 5%(ANOVA)	1456	0.00	0.8	0.28±0.05	0.2
kNN on 50(ANOVA)	50	0.01	0.8	0.07±0.02	0.0
kNN on SMLR(lm=10) non-0	42	0.00	2.5	0.12±0.02	0.0
LinSVM on 5%(SVM)	1456	0.00	23.1	0.18±0.04	0.1
LinSVM on 50(SVM)	50	0.00	22.7	**0.03**±0.02	0.0
LinSVM on 5%(ANOVA)	1456	0.00	1.6	0.13±0.04	0.1
LinSVM on 50(ANOVA)	50	0.00	0.8	0.09±0.03	0.0
LinSVM on SMLR(lm=1) non-0	92	0.00	5.3	0.08±0.02	0.0
LinSVM on SMLR(lm=10) non-0	42	0.00	2.5	0.12±0.03	0.0
LinSVM+RFE(N-Fold)	4587	0.00	2010.0	0.12±0.03	3.4
LinSVM+RFE(OddEven)	42	0.09	260.9	0.24±0.04	0.0

CHAPTER 6

SUMMARY

There is no such thing as failure, only results, with some more successful than others

– Jeff Keller *Attitude is Everything, Inc.*

6.1 Conclusion

This dissertation presented a unified methodology toward the analysis of neural data from different or even multiple neural data modalities. The demonstrated approach is generic enough to be used across different levels of investigation of brain function while enforcing reliable decoding performance. The approach ensures unbiased testing of results and allows conclusions to be drawn about neural processes of interest. Furthermore, analysis of constructed decoders provides the basis for addressing *identification*, and the assessment of *specificity* in a way not possible with conventional approaches. The results obtained with this approach revealed the components of the signals which would not be detected using conventional approaches. Decoding of extracellular recordings allowed to validate presence of stimuli-specific information at recording cite, as well as to assign stimulus specific labels to the principal functional units (*i.e.*, neurons). Investigation of EEG data decoder sensitivities allowed to detect late stimulus-sensitive component of the ERP. Analysis of object specific visual processing allowed to support the hypothesis of distributed neural codes and the absence of the FFA, uniquely involved in human face recognition.

Furthermore, mapping of the neural data from one imaging modality into another allowed the segregation of neural data components present within the empirical data, and

provided the means for performing a variety of additional analysis that ranged from data filtering to spatial localization of the neural processes.

6.2 Contributions Beyond The Dissertation

The signal decoding framework presented in this dissertation is applicable to a number of sub-fields in neuroscience, from extracellular recordings to cross-modal data analysis. The value of this approach has been demonstrated with a number of presentations at neuroscience conferences and several articles in peer-reviewed journals [22, 23, 28].

This framework has also stimulated a variety of collaborative efforts across various aspects of human brain function, such as causal inference [226] and large scale decoding [227]. To facilitate the application of the analysis approach presented here, a Python-based toolbox, PyMVPA, has been developed in a joint effort with Michael Hanke (Germany), Per Sederberg (USA), and Emanuele Olivetti (Italy) [22, 23]. As a software product PyMVPA has significantly matured over the last few years and has already been used by a number of research groups and individual researchers around the globe, including those working at MIT, UCLA, and a number of other universities and research institutes. PyMVPA has been developed as part of ExpPsy (*Experimental Psychology*[1]) project within the *Debian GNU/Linux*[2] operating system, of which I *am*[3]. one of the proud developers. The ExpPsy project aims to provide the best, free, and open-source research environment for psychologists and neuroscientists. Various software products (*e.g.*, PyEPL, FSL, PyNIFTI, OpenMEEG, *etc.*) germane to research of the brain were either developed or packaged and maintained within ExpPsy. Due to the terrific packaging system of Debian, ExpPsy made these software products readily available using a single click or simple command line operation. These packages are

[1] http://alioth.debian.org/projects/pkg-exppsy/
[2] http://www.debian.org
[3] http://qa.debian.org/developer.php?login=debian@onerussian.com

available not only to the thousands of Debian users, but to the users of Debian-derived distributions (*e.g.*, Ubuntu) on a variety of platforms and architectures.

6.3 Future Work

The results that were presented here have demonstrated the power of statistical learning methods applied toward the analysis of neural data. Nevertheless, there are a number of ways in which the suggested framework can be improved and extended.

Although the results that have been presented here were based on constructing a linear relation between input and output data, the proposed framework is in no way limited to linear mapping. Non-linear mapping is of particular value for cross-modal data mapping which was presented in the Chapter 4. The main obstacle in adopting non-linear methods is the necessity to explore the larger parameter space of more flexible non-linear models. It is computationally demanding and often requires ad-hoc optimization. Although some ability to do model selection is already present in PyMVPA, the capacity for transparent model selection within PyMVPA is not yet implemented. Development of such a feature would rapidly facilitate the exploration of non-linear learners in a variety of applications.

At the moment, cross-modal mapping presented in the Chapter 4 makes no use of the EEG forward model. However, although the incorporation of a known forward model might significantly boost reconstruction performance, it is not yet implemented.

Although multi-variate methods, and decoders in particular, make use of the covariance and causality structure embedded in the data, the results presented here do not actually reveal any such *effective connectivity*. Nevertheless, preliminary results (not yet published) demonstrate the viability of using statistical learning regressions and associated sensitivity estimates to discover functional connectivity among different brain regions. Further development of such methods would be another goal for future research.

Although the proposed framework offers considerable value to researchers studying the brain, decoders and statistical learning do not offer a panacea for all research problems in neuroscience. The success of this methodology requires that researchers consider well both the research question and the experimental design. Any particular investigation of brain function, which might seem obvious at first sight, needs to be approached with considerable caution and domain knowledge. Unfortunately, direct applications of classifiers [228] on psychologically complex problems, such as deception detection [229], can be overly *naive* to be practical [230]. Thus, another goal to be pursued is the development of a guide for researchers that will facilitate the formulation of research questions and experimental design that maximizes the value of this framework.

APPENDIX A

FORWARD MODELING OF EEG AND MEG SIGNALS

The analysis of E/MEG signals often relies on the solution of two related problems. The *forward problem* concerns the calculation of scalp potentials (EEG) or magnetic fields near the scalp (MEG) given the neuronal currents in the brain, whereas the *inverse problem* involves estimating neuronal currents from the observed E/MEG data. The difficulty of solving the forward problem is reflected in the diversity of approaches that have been tried (see [231] for an overview and unified analysis of different methods).

The basic question posed by both the inverse and forward problems is how to model any neuronal activation so that the source of the electromagnetic field can be mapped onto the observed E/MEG signal. Assuming that localized and synchronized primary currents are the generators of the observed E/MEG signals, the most successful approach is to model the i-th source with a simple Equivalent Current Dipole (ECD) $\mathbf{q}_i$ [232], uniquely defined by three factors: location represented by the vector $\mathbf{r}_i$, strength q_i, and orientation coefficients θ_i. The orientation coefficient is defined by projections of the vector $\mathbf{q}_i$ into L orthogonal Cartesian axes: $\theta_i = \mathbf{q}_i/q_i$. However, the orientation coefficient may be expressed by projections in two axes in the case of a MEG spherical model where the silent radial to the skull component has been removed, or even, just in a single axis if normality to the cortical surface is assumed. The ECD model made it possible to derive a tractable physical model linking neuronal activation and observed E/MEG signals. In case of K simultaneously active sources at time t the observed E/MEG signal at the sensor $\mathbf{x}_j$ positioned at $\mathbf{p}_j$ can be modeled as

$$\hat{\mathbf{x}}_j(\mathbf{r}_i, \mathbf{q}_i, t) = \sum_{i}^{K} \mathcal{G}(\mathbf{r}_i(t), \mathbf{p}_j) \cdot \mathbf{q}_i(t) + \epsilon, \tag{A.1}$$

where $\mathcal{G}$ is a *lead field* function which relates the i-th dipole and the potential (EEG) or magnetic field (MEG) observed at the j-th sensor; and ϵ is the sensor noise. In the given formulation, function $\mathcal{G}(\mathbf{r}_i(t), \mathbf{p}_j)$ returns a vector, where each element corresponds to the lead coefficient at the location $\mathbf{p}_j$ generated by a unit-strength dipole at position $\mathbf{r}_i(t)$ with the same orientation as the corresponding projection axis of θ_i. The inner-product between the returned vector and dipole strength projections on the same coordinate axes yields a j-th sensor the measurement generated by the i-th dipole.

The forward model (A.1) can be solved at substantial computational expense using available numerical methods [233] in combination with realistic structural information obtained from the MRI data. This high computational cost is acceptable when the forward model has to be computed once per subject and for a fixed number of dipole locations, but it can be prohibitive for dipole fitting, which requires a recomputation of the forward model for each step of non-linear optimization. For this reason, rough approximations of the head geometry and structure are often used: *e.g.*, best-fit single sphere model which has a direct analytical solution [234] or the multiple spheres model to accommodate for the difference in conductivity parameters across different tissues. Recently proposed MEG forward modeling methods for realistic isotropic volume conductors [235, 236] are more accurate and faster than Boundary Elements Method (BEM), and hence may be useful substitutes for both crude analytical methods and computationally intensive finite-element numeric approximations.

APPENDIX B

EEG AND MEG INVERSE PROBLEM

B.1 Equivalent Current Dipole Models

The E/MEG inverse problem is very challenging (see [237, 238] for an overview of methods.) First, it relies on the solution of the forward problem, which can be computationally expensive, especially in the case of realistic head modeling. Second, the lead-field function $\mathcal{G}$ from (A.1) is non-linear in $\mathbf{r}_i$, so that the forward model depends non-linearly on the locations of activations. It is because of this nonlinearity that the inverse problem is generally treated by non-linear optimization methods, which can lead to solutions being trapped in local minima. In case of Gaussian sensor noise, the best estimator for the reconstruction quality of the signal is the squared error between the obtained and modeled E/MEG data:

$$\mathcal{E}(\mathbf{r},\mathbf{q}) = \sum_{i}^{K}\sum_{t=t_1}^{t_2}\sum_{j}^{M}\big(\mathbf{x}_j(t) - \hat{\mathbf{x}}_j(\mathbf{r}_i,\mathbf{q}_i,t)\big)^2 + \lambda\mathcal{C}(\mathbf{r},\mathbf{q}), \tag{B.1}$$

where $\mathcal{C}(\mathbf{r},\mathbf{q}) > 0$ is often introduced to regularize the solution, *i.e.*, to obtain the desired features of the estimated signal (*e.g.*, smoothness in time, or in space, lowest energy or dispersion), and $\lambda > 0$ is used to vary the trade-off between the goodness of fit and the regularization term.

This least-squares model can be applied to the individual time-points ($t_1 = t_2$) ("moving dipole" model) or to a block ($t_1 < t_2$) of data points. If the sources are assumed to stay constant during the block (t_1,t_2), then the solution with time constant $\mathbf{q}_i(t) = \mathbf{q}_i$ is the target.

Other features derived from the data besides pure E/MEG signals as the argument $\mathbf{x}$ of (A.1) and (B.1) are often used: *e.g.*, ERP/ERF waveforms which represent averaged E/MEG signals across multiple trials, mean map in the case of stable

potential/field topography during some period of time, or signal frequency components to localize the sources of the oscillations of interest.

Depending on the treatment of (B.1), the inverse problem can be presented in a couple of different ways. The brute-force minimization of (B.1) in respect to both parameters $\mathbf{r}$ and $\mathbf{q}$, and the consideration of different K neuronal sources, is generally called *ECD fitting*. Because of non-linear optimization, this approach works only for cases where there is a relatively small number of sources K, and therefore the inverse problem formulation is over-determined, *i.e.*, (A.1) has no exact solution ($\mathcal{E}(\mathbf{r}, \mathbf{q}) > 0$). If fixed time locations of the target dipoles can be assumed, the search space of non-linear optimization is reduced and the optimization can be split into two steps: (a) non-linear optimization to find locations of the dipoles, and then (b) analysis to determine the strength of the dipoles. This assumption constitutes the so-called *spatiotemporal ECD model*.

Two other frameworks have been suggested as means of avoiding the pitfalls associated with non-linear optimization: Distributed ECD (DECD) and beamforming. These two approaches are presented in detail in the next sections.

B.2 Linear Inverse Methods: Distributed ECD

In case of multiple simultaneously active sources, an alternative to solving the inverse problem by ECD fitting is a distributed source model. The label Distributed ECD (DECD) will be used further in the text to refer to this type of model. The DECD is based on a spatial sampling of the brain volume and distributing the dipoles across all plausible and spatially small areas, which could be a source of neuronal activation. In such cases, fixed locations ($\mathbf{r}_i$) are available for each source/dipole, removing the necessity of non-linear optimization as in the case of the ECD fitting. The forward model (A.1) can be presented for a noiseless case in the matrix form

$$\mathbf{X} = \mathbf{G}\mathbf{Q}, \tag{B.2}$$

where $\mathbf{G}$, $M \times LN$ *lead field* matrix, is assumed to be static in time. The j, i-th entry of $\mathbf{G}$ describes how much a sensor j is influenced by a dipole i, where j varies over all sensors while i varies over every possible source, or to be more specific, every axis-aligned component of every possible source: $g_{j\bar{\imath}} = \mathcal{G}(\mathbf{r}_i, \mathbf{p}_j)$. The vector $\bar{\imath}$ contains indices of L such projections, *i.e.*, $\bar{\imath} = [i, i + N, i + 2N]$ when $L = 3$, and $\bar{\imath} = i$ when the dipole has a fixed known orientation. Using this notation, $\mathbf{G}_{\cdot,\bar{\imath}}$ corresponds to the lead matrix for a single dipole $\mathbf{q}_i$. The $M \times T$ matrix $\mathbf{X}$ holds the E/MEG data, while the $LN \times T$ matrix $\mathbf{Q}$ (note that $\mathbf{Q}_{\bar{\imath}t} = \mathbf{q}_i(t)$) corresponds to the projections of the ECD's moment onto L orthogonal axes.

The solution of (B.2) relies on finding an inverse $\mathbf{G}^+$ of the matrix $\mathbf{G}$ to express the estimate $\hat{\mathbf{Q}}$ in terms of $\mathbf{X}$

$$\hat{\mathbf{Q}} = \mathbf{G}^+ \mathbf{X}, \tag{B.3}$$

and will produce a linear map $\mathbf{X} \longmapsto \hat{\mathbf{Q}}$. Other than being computationally convenient, there is no much reason to take this approach. The task is to minimize the error function (B.1), which can be generalized by the weighting of the data to account for the sensor noise and its covariance structure:

$$\mathcal{E}(\mathbf{Q}) = \mathrm{tr}\big((\mathbf{X} - \mathbf{G}\mathbf{Q})^\top \mathbf{W}_\mathbf{X}^{-1} (\mathbf{X} - \mathbf{G}\mathbf{Q})\big), \tag{B.4}$$

where $\mathbf{W}_\mathbf{X}^{-1}$ is a weighting matrix in sensor space.

A zero-mean Gaussian signal can be characterized by the single covariance matrix $\mathbf{C}_\epsilon$. In case of a non-singular $\mathbf{C}_\epsilon$ the most simple weighting scheme $\mathbf{W}_\mathbf{X} = \mathbf{C}_\epsilon$ can be used to account for non-uniform and possibly correlated sensor noise.

Such a brute-force approach solves some problems of ECD modeling, specifically the requirement for a non-linear optimization, but, unfortunately, it introduces another problem: the linear system (B.2) is ill-posed and under-determined because the number of sampled possible source locations is much higher than the dimensionality of the input

data space (which cannot exceed the number of sensors), *i.e.*, $N \gg M$. Thus, there is an infinite number of solutions for the linear system because any combination of terms from the null space of $\mathbf{G}$ will satisfy equation (B.3) and fit the sensor noise perfectly. In other words, many different arrangements of the sources of neural activation within the brain can produce any given MEG or EEG map. To overcome such ambiguity, a regularization term is introduced into the error measure

$$\mathcal{E}_r(\mathbf{Q}) = \mathcal{E}(\mathbf{Q}) + \lambda \mathcal{C}(\mathbf{Q}), \tag{B.5}$$

where $\lambda \geq 0$ controls the trade-off between the goodness of fit and the regularization term $\mathcal{C}(\mathbf{Q})$.

The equation (B.5) can have different interpretations depending on the approach used to derive it and the meaning given to the regularization term $\mathcal{C}(\mathbf{Q})$. All of the following methods provide the same result under specific conditions [238, 239]: Bayesian methodology to maximize the posterior $p(\mathbf{Q}|\mathbf{X})$ assuming Gaussian prior on $\mathbf{Q}$ [240], Wiener estimator with proper $\mathbf{C}_\epsilon$ and $\mathbf{C}_S$, Tikhonov regularization to trade-off the goodness of fit (B.4) and the regularization term $\mathcal{C}(\mathbf{Q}) = \mathrm{tr}(\mathbf{Q}^\top \mathbf{W}_\mathbf{Q}^{-1} \mathbf{Q})$ which attempts to find the solution with weighted by $\mathbf{W}_\mathbf{Q}^{-1}$ minimal 2nd norm, which is usually called *Weighted minimum norm* (WMN). All the frameworks lead to the solution of the next general form

$$\mathbf{G}^+ = (\mathbf{G}^\top \mathbf{W}_\mathbf{X}^{-1} \mathbf{G} + \lambda \mathbf{W}_\mathbf{Q}^{-1})^{-1} \mathbf{G}^\top \mathbf{W}_\mathbf{X}^{-1}. \tag{B.6}$$

If and only if $\mathbf{W}_\mathbf{Q}$ and $\mathbf{W}_\mathbf{X}$ are positive definite [241] (B.6) is equivalent to

$$\mathbf{G}^+ = \mathbf{W}_\mathbf{Q} \mathbf{G}^\top (\mathbf{G} \mathbf{W}_\mathbf{Q} \mathbf{G}^\top + \lambda \mathbf{W}_\mathbf{X})^{-1}. \tag{B.7}$$

In case when viable prior information about the source distribution is available $\mathbf{Q}_p$, it is easy to account for it by minimizing the deviation of the solution not from $\mathbf{0}$ (which constitutes the minimal 2nd norm solution $\mathbf{G}^+$), but from the prior $\mathbf{Q}_p$, *i.e.*, $\mathcal{C}(\mathbf{Q}) =$

$\mathrm{tr}\big((\mathbf{Q}-\mathbf{Q}_p)^\top \mathbf{W}_\mathbf{Q}^{-1}(\mathbf{Q}-\mathbf{Q}_p)\big)$. Then (B.5) will be minimized at

$$\hat{\mathbf{Q}} = \mathbf{G}^+\mathbf{X} + (\mathbf{I}-\mathbf{G}^+\mathbf{G})\mathbf{Q}_p = \mathbf{Q}_p + \mathbf{G}^+(\mathbf{X}-\mathbf{G}\mathbf{Q}_p). \tag{B.8}$$

For the noiseless case, with a weighted L_2-norm regularizer, the Moore-Penrose pseudo-inverse gives the inverse $\mathbf{G}^+ = \mathbf{G}^\dagger$ by avoiding the null space projections of $\mathbf{G}$ in the solution, thus providing a unique solution with a minimal second norm $\mathbf{G}^\dagger = \mathbf{W}_\mathbf{Q}\mathbf{G}^\top(\mathbf{G}\mathbf{W}_\mathbf{Q}\mathbf{G}^\top)^{-1}$.

Taking $\mathbf{W}_\mathbf{Q} = \mathbf{I}_N, \mathbf{W}_\mathbf{X} = \mathbf{I}_M$ and $\mathbf{Q}_p = \mathbf{0}$ constitutes the simplest regularized minimum norm solution (Tikhonov regularization). Classically, λ is found using cross-validation [242] or L-curve [243] techniques, to decide how much of the noise power should be brought into the solution. Phillips et al. [244] suggested iterative method ReML where the conditional expectation of the source distribution and the regularization parameters are estimated jointly. ReML was shown [245] to perform better than a regular WMN in the case of invalid location priors. Additional constraints can be added to impose an additional regularization: for instance temporal smoothness [246].

As presented in (B.7), $\mathbf{G}^+$ can account for different features of the source or data space by incorporating them correspondingly into $\mathbf{W}_\mathbf{Q}$ and $\mathbf{W}_\mathbf{X}$. Next data-driven features are commonly used in EMSI

- $\mathbf{W}_\mathbf{X} = \mathbf{C}_\epsilon$ accounts for any possible noise covariance structure or, if $\mathbf{C}_\epsilon$ is diagonal, will scale the error terms according to the noise level of each sensor;
- $\mathbf{W}_\mathbf{Q} = \mathbf{W}_{\mathbf{C}_S} = \mathbf{C}_S$ accounts for prior knowledge of the sources covariance structure.

$\mathbf{W}_\mathbf{Q}$ can also account for different spatial features

- $\mathbf{W}_\mathbf{Q} = \mathbf{W}_\mathrm{n} = \big(\mathrm{diag}\,(\mathbf{G}^\top\mathbf{G})\big)^{-p}$ normalizes the columns of the matrix $\mathbf{G}$ to account for deep sources by penalizing voxels too close to the sensors [247, 248]. Weighting degree $p \in [0.6 \ldots 0.8]$ was shown to improve mean localization accuracy [249];

- $\mathbf{W_Q} = \mathbf{W}_{\text{gm}}$, where the i-th diagonal element incorporates the gray matter content in the area of the i-th dipole [250], *i.e.*, the probability of having a large population of neurons capable of creating the detected E/MEG signal;

- $\mathbf{W_Q} = (\mathbf{W}_a{}^{\top}\mathbf{W}_a)^{-1}$, where rows of $\mathbf{W}_a$ represent averaging coefficients for each source [251]. So far only geometrical [252] or biophysical averaging matrices [241] were suggested;

- $\mathbf{W_Q}$ incorporates the first-order spatial derivative of the image [253] or Laplacian form [254].

Features defined by the diagonal matrices (*e.g.*, $\mathbf{W}_{\text{n}}$ and $\mathbf{W}_{\text{gm}}$) can be combined through the simple matrix product. An alternative approach is to present $\mathbf{W_Q}$ in terms of a linear basis set of the individual $\mathbf{W_Q}$ factors, *i.e.*, $\mathbf{W_Q} = \mu_1\mathbf{W}_{\text{n}} + \mu_2\mathbf{W}_{\text{gm}} + \cdots$, with later optimization of μ_i via the EM algorithm [250].

To better condition the under-determined linear inverse problem (B.3), Phillips et al. [250] suggested to perform the inverse operation in the space of the largest eigenvectors of the $\mathbf{W_Q}$. Such preprocessing can also be done in the temporal domain, when a similar sub-space selection is performed using prior temporal covariance matrix, thus effectively selecting the frequency power spectrum of the estimated sources.

Careful selection of the described features of data and source spaces helps to improve the fidelity of the DECD solution. Nevertheless, the inherent ambiguity of the inverse solution precludes achieving a high degree of localization precision. It is for this reason that additional spatial information about the source space, readily available from other functional modalities such as fMRI and PET, can help to condition the DECD solution (Section 4.3.4).

B.3 Beamforming

Beamforming (sometimes called a spatial filter or a virtual sensor) is another way to solve the inverse problem, which actually does not directly minimize (B.1). A beamformer attempts to find a linear combination of the input data $\hat{\mathbf{q}}_i = \mathbf{F}^i\mathbf{x}$, which represents the neuronal activity of each dipole $\mathbf{q}_i$ in the best possible way one at a given time. As in DECD methods, the search space is sampled, but, in contrast to the DECD approach, the beamformer does not try to fit all the observed data at once.

The linearly constrained minimum variance (LCMV) beamformer [255] looks for a spatial filter defined as $\mathbf{F}^i$ of size $M \times L$ minimizing the output energy ${\mathbf{F}^i}^\top \mathbf{C}_X \mathbf{F}^i$ under the constraint that only $\mathbf{q}_i$ is active at that time, *i.e.*, that there is no attenuation of the signal of interest: $\mathbf{F}^k\mathbf{G}_{\cdot,\bar{\imath}} = \delta_{ki}\mathbf{I}_L$, where the Kronecker delta $\delta_{ki} = 1$ only if $k = i$ and 0 otherwise. Because the beamforming filter $\mathbf{F}^i$ for the i-th dipole is defined independently from the other possible dipoles, index i will be dropped from the derived results for the clarity of presentation.

The constrained minimization, solved using Lagrange multipliers, yields

$$\mathbf{F} = ({\mathbf{G}_{\cdot,\bar{\imath}}}^\top \mathbf{C}_X^{-1} \mathbf{G}_{\cdot,\bar{\imath}})^{-1} {\mathbf{G}_{\cdot,\bar{\imath}}}^\top \mathbf{C}_X^{-1} \tag{B.9}$$

This solution is equivalent to (B.6), when applied to a single dipole with the regularization term omitted. Source localization is performed using (B.9) to compute the variance of every dipole $\mathbf{q}$, which, in the case of uncorrelated dipole moments, is

$$\nu_{\mathbf{q}} = \mathrm{tr}\big((\mathbf{G}_{\cdot,\bar{\imath}}{}^\top \mathbf{C}_X^{-1} \mathbf{G}_{\cdot,\bar{\imath}})^{-1}\big). \tag{B.10}$$

The noise-sensitivity of (B.10) can be reduced by using the noise variance of each dipole as normalizing factor $\nu_\epsilon = \mathrm{tr}\big((\mathbf{G}_{\cdot,\bar{\imath}}{}^\top \mathbf{C}_\epsilon^{-1} \mathbf{G}_{\cdot,\bar{\imath}})^{-1}\big)$. This produces the so-called *neural activity index*

$$z = \frac{\nu_{\mathbf{q}}}{\nu_\epsilon}. \tag{B.11}$$

An alternative beamformer, *synthetic aperture magnetometry* or SAM [256], is similar to the LCMV if the orientation of the dipole is defined, but it is quite different in the case of a dipole with an arbitrary orientation. A vector of lead coefficients $\mathbf{g}_i(\theta)$ is defined as a function of the dipole orientation. This returns a single vector for the orientation θ of the i-th dipole, as opposed to the earlier formulation in which the L columns of $\mathbf{G}_{\cdot,\bar{\imath}}$ played a similar role. With this new formulation, the spatial filter is constructed

$$\mathbf{f}(\theta) = \frac{1}{\mathbf{g}_i(\theta)^\top \mathbf{C}_X^{-1} \mathbf{g}_i(\theta)} \mathbf{g}_i(\theta)^\top (\mathbf{C}_X + \lambda \mathbf{C}_\epsilon)^{-1} \tag{B.12}$$

which, under standard assumptions, is an optimal linear estimator of the time course of the i-th dipole. The variance of the dipole, accordingly, is also a function of θ, specifically $\nu_{\mathbf{q}}(\theta) = 1/\big(\mathbf{g}_i(\theta)^\top \mathbf{C}_X^{-1} \mathbf{g}_i(\theta)\big)$. To compute the neuronal activity index the original SAM formulation uses a slightly different normalization factor $\nu_\epsilon(\theta) = \mathbf{f}(\theta)^\top \mathbf{C}_\epsilon \mathbf{f}(\theta)$, which yields a different result if the noise variance in $\mathbf{C}_\epsilon$ is not equal across the sensors.

The unknown value of θ is found via a non-linear optimization of the neuronal activity index for the dipole:

$$\theta = \arg\max_{\vartheta} \frac{\nu_{\mathbf{q}}(\vartheta)}{\nu_\epsilon(\vartheta)}.$$

Despite the pitfalls of non-linear optimization, SAM filtering provides a higher SNR to LCMV by bringing less than half of the noise power into the solution. In addition, SAM filtering results in sharper peaks of the distribution of neuronal activity index over the volume [257].

Having computed $\nu_{\mathbf{q}}$ and ν_ϵ using SAM or LCMV for the two experimental conditions: passive (p) and active (a), it is possible to compute a pseudo-t value $\hat{t}$ for each location across the two conditions

$$\hat{t} = \frac{\nu_{\mathbf{q}}^{(a)} - \nu_{\mathbf{q}}^{(p)}}{\nu_\epsilon^{(a)} + \nu_\epsilon^{(p)}}. \tag{B.13}$$

Such an approach provides the possibility of considering experimental design in the analysis of E/MEG localization.

Unlike ECD, beamforming does not require prior knowledge of the number of sources, which is a non-trivial problem on its own [258]. Nor beamforming does its search for a solution in an underdetermined linear system as does DECD. For these reasons, beamforming remains the favorite method of many researchers in EMSI and has been suggested for use in the integrative analysis of E/MEG and fMRI which is covered in Section 4.3.5.

APPENDIX C

DETAILS OF THE MULTIMODAL DATA ANALYSIS

This appendix provides additional information about multimodal analysis of the data, which is presented in Section 4.4.

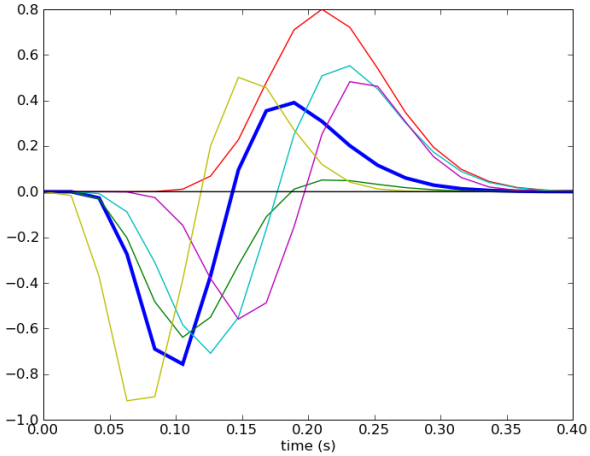


Figure C.1 LFP responses used for simulation. Blue LFP is the one used for modeling event-related activity, while all the others (including the blue one as well) were arbitrarily assigned to voxels to generate simulated LFP response.

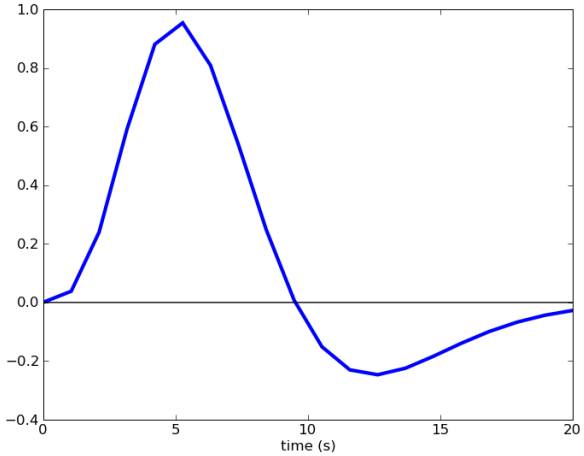


Figure C.2 HRF response used for fMRI data simulation.

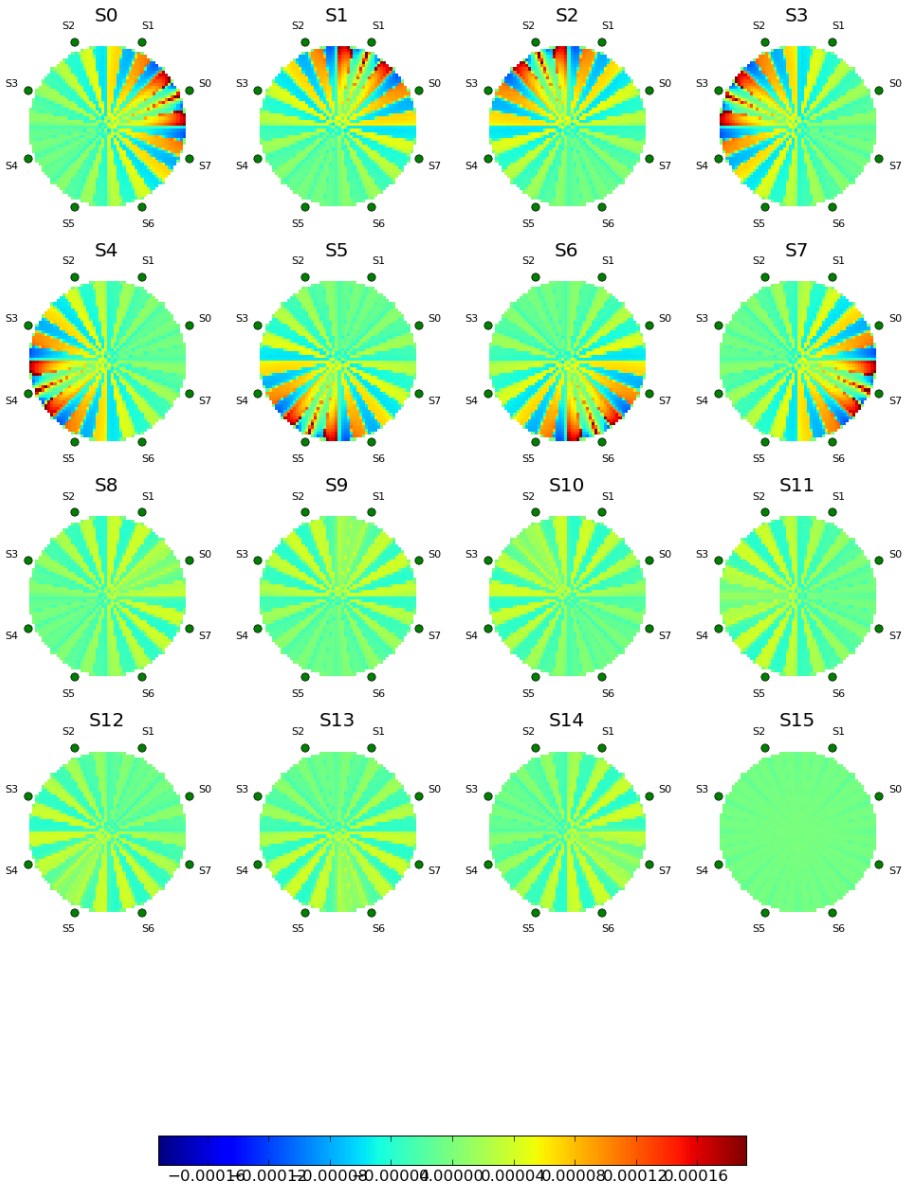


Figure C.3 Gain matrix for EEG signal simulation.

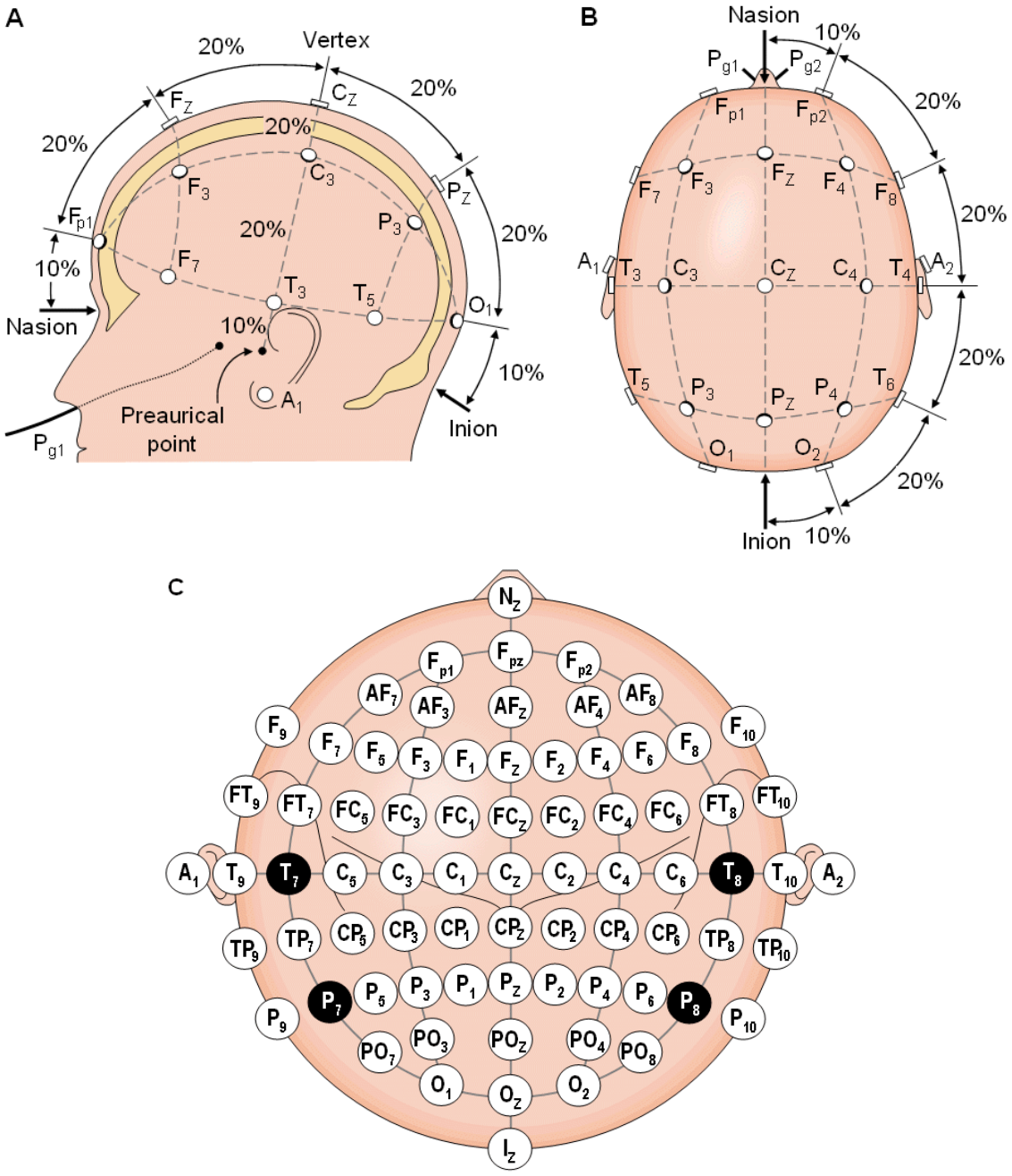


Figure C.4 The international 10-20 system seen from (A) left and (B) above the head. A = Ear lobe, C = central, Pg = nasopharyngeal, P = parietal, F = frontal, Fp = frontal polar, O = occipital. (C) Location and nomenclature of the intermediate 10% electrodes, as standardized by the American Electroencephalographic Society. (From [259])

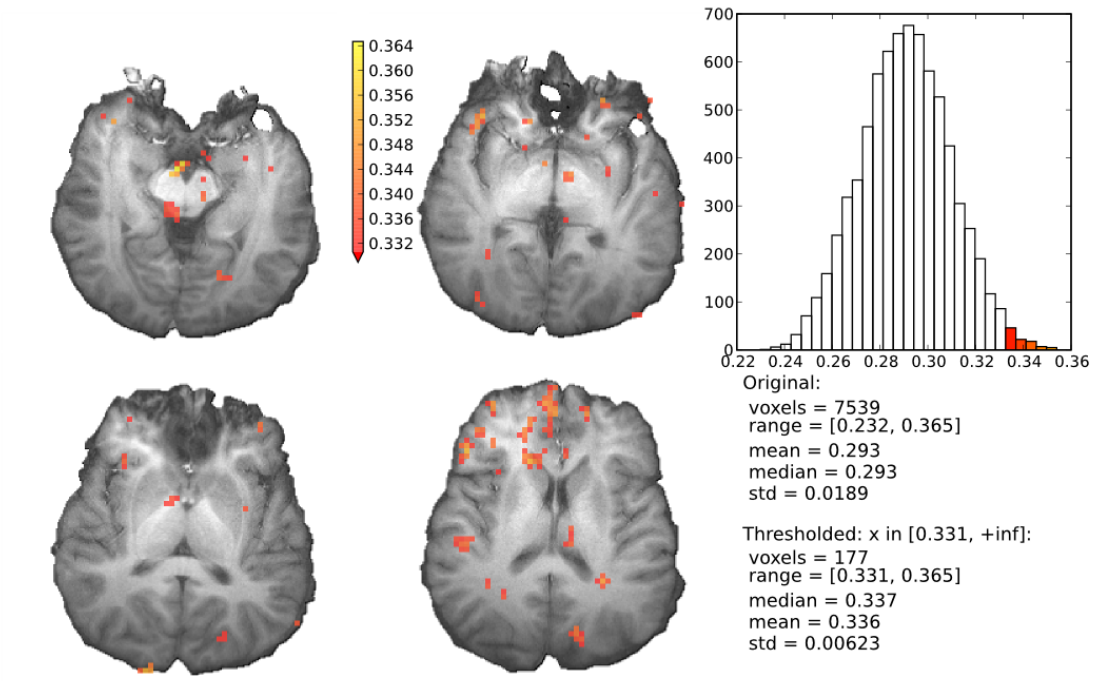


Figure C.5 Aggregate sensitivity to the frequencies in the δ_1-band range of frequencies.

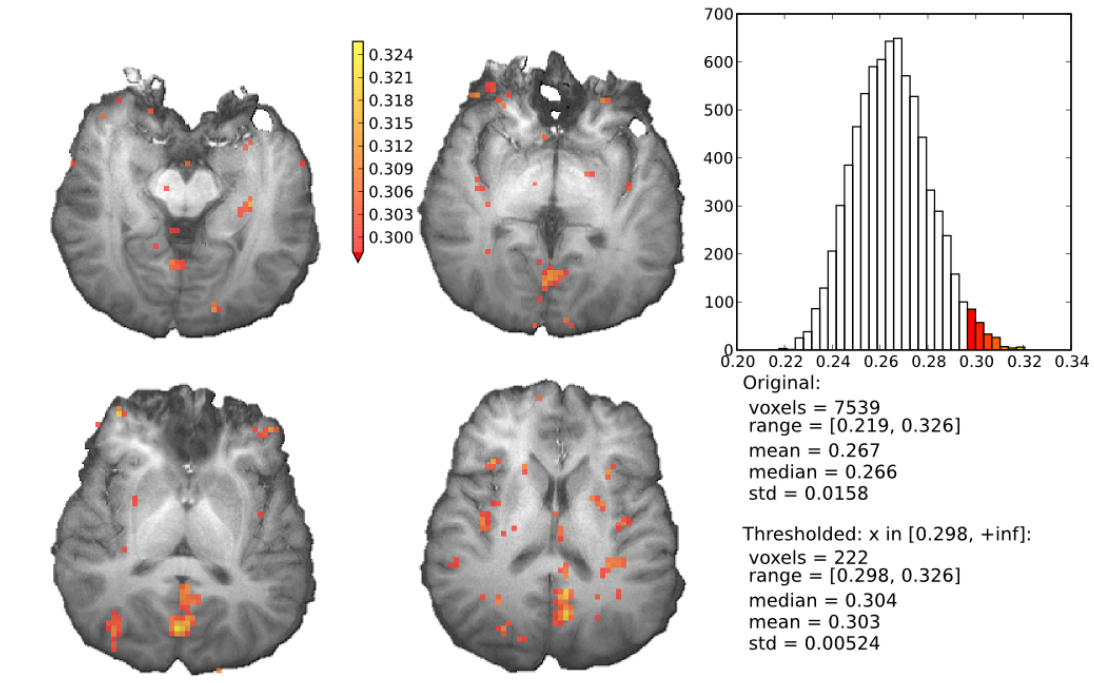


Figure C.6 Aggregate sensitivity to the frequencies in the δ_2-band range of frequencies.

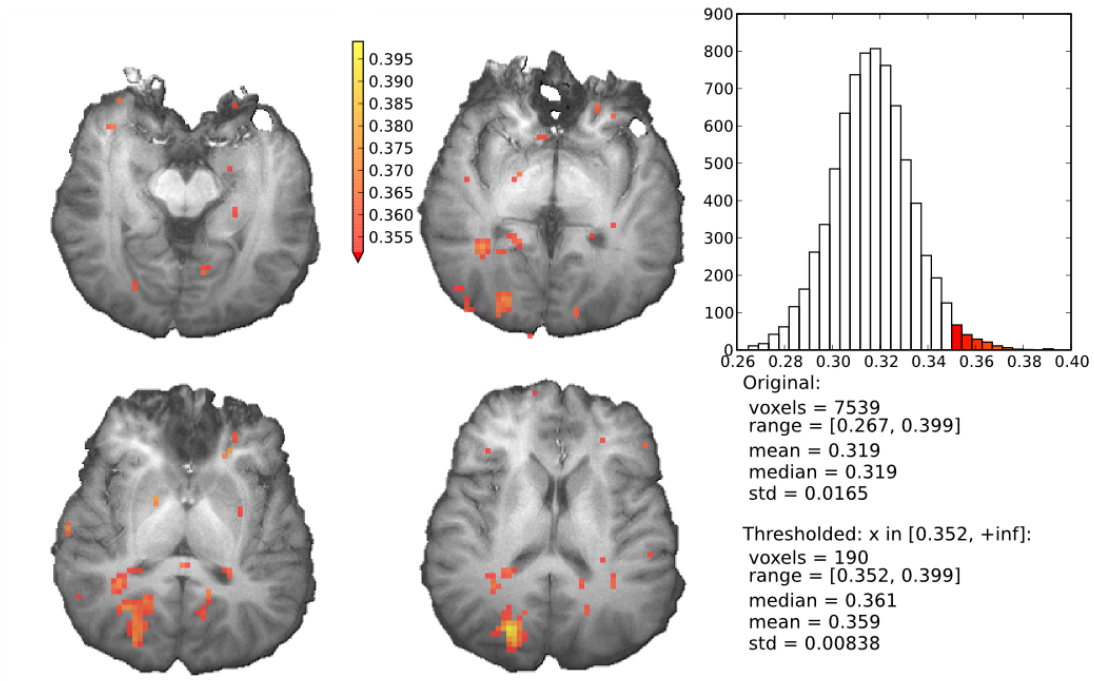


Figure C.7 Aggregate sensitivity to the frequencies in the θ-band range of frequencies.

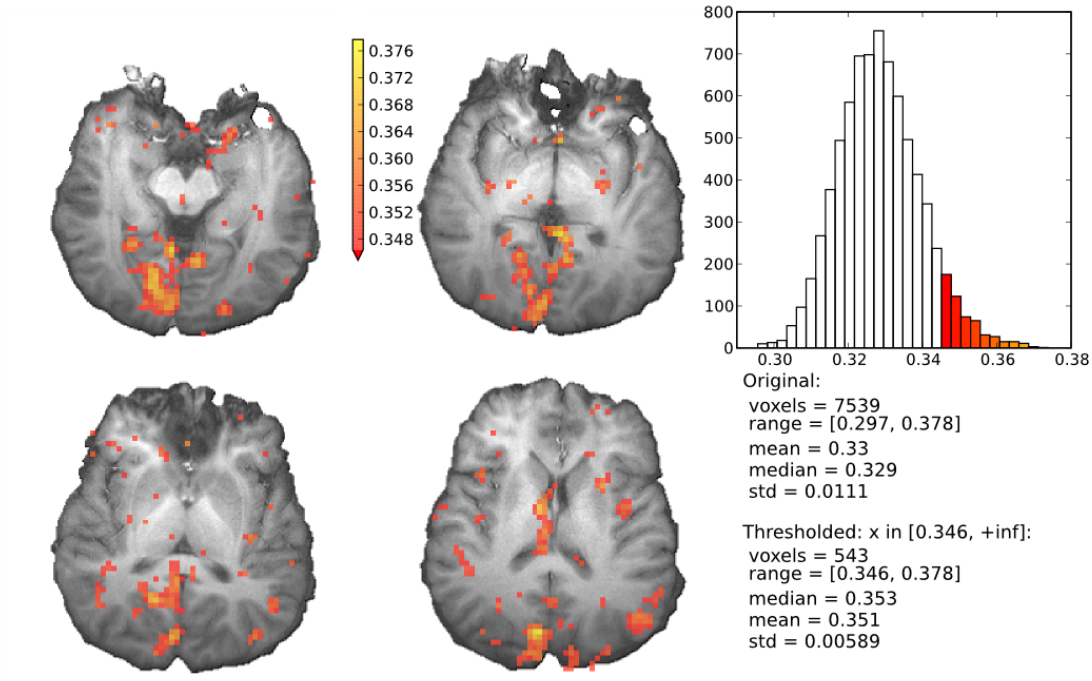


Figure C.8 Aggregate sensitivity to the signal of CP1 EEG channel (see Figure C.4 for the typical location of the sensor in respect to the brain).

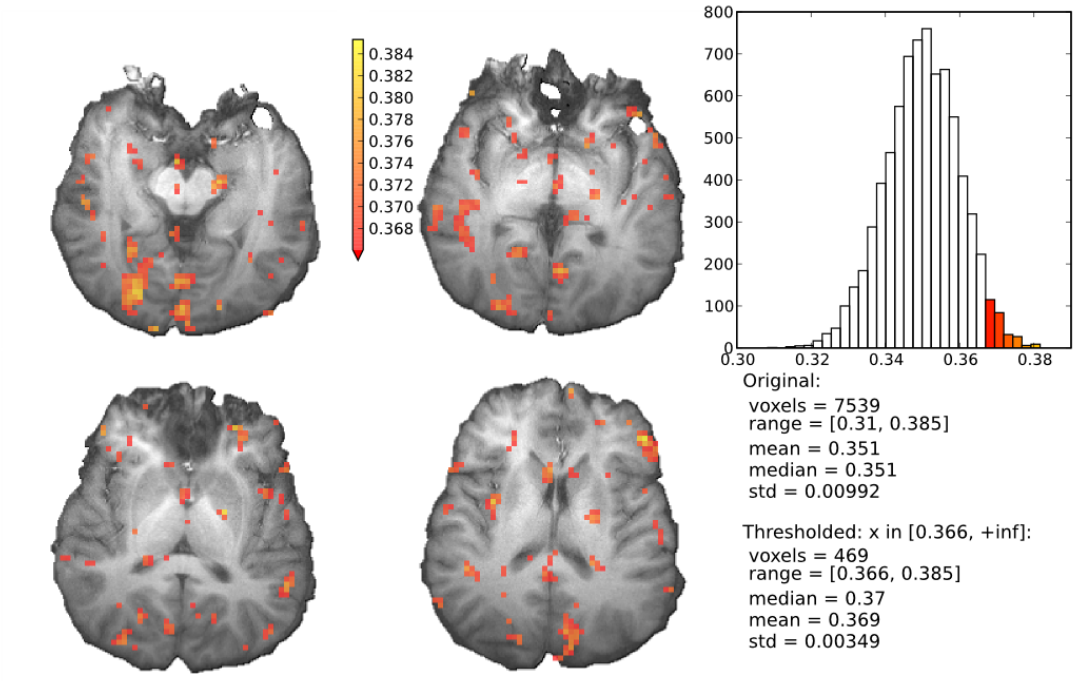


Figure C.9 Aggregate sensitivity to the signal of CP2 EEG channel (see Figure C.4 for the typical location of the sensor in respect to the brain).

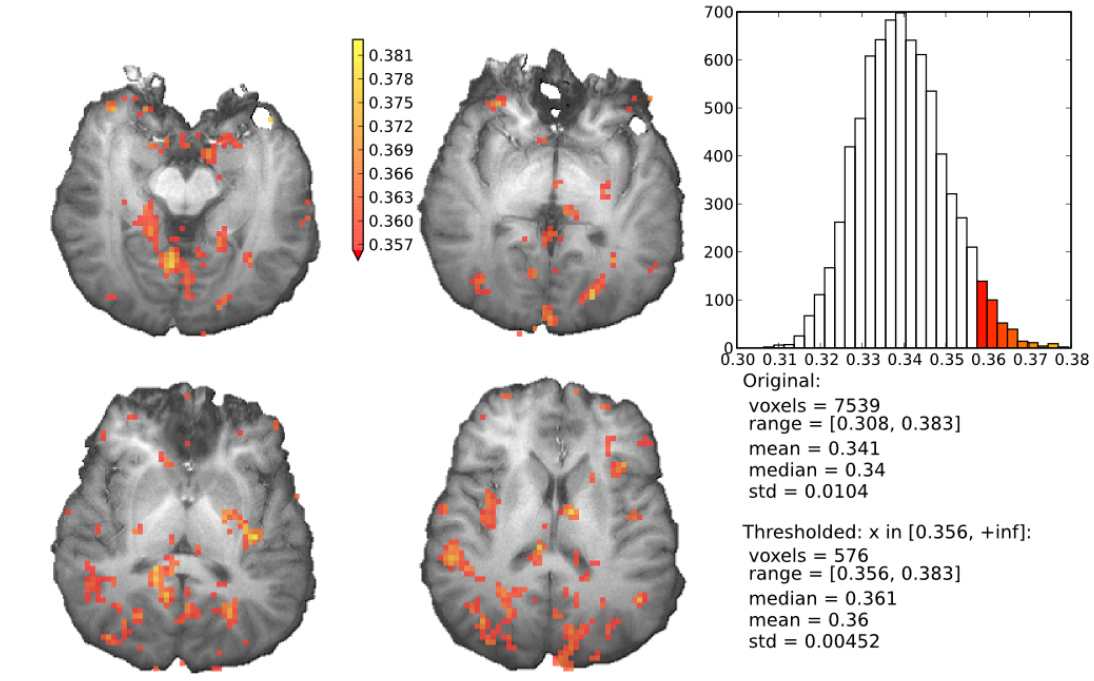


Figure C.10 Aggregate sensitivity to the signal of Pz EEG channel (see Figure C.4 for the typical location of the sensor in respect to the brain).

APPENDIX D

FREE OPEN SOURCE SOFTWARE GERMANE TO THE ANALYSIS OF NEUROIMAGING DATA

Table D.1 lists a handful of FOSS projects relevant to the analysis of neuroimaging analysis, whenever Table D.2 lists projects, either written in Python or providing Python bindings, which are germane to acquiring or processing neural information datasets using machine learning (ML) methods. The last column in Table D.2 indicates whether PyMVPA internally uses a particular project or provides public interfaces to it.

Table D.1 Free Software Germane to the Analysis of Neuroimaging Data.

Package	Forward				Inverse		MRI†				fMRI	Envir.		
	EEG			MEG										
	FEM	BEM	Spherical	FEM	BEM	Spherical	Beamforming	DECD	ECD	E/MEG Regist.	Tessellation	Scalp Segm.	Skull Segm.	Brain Segm.		MS Windows	Mac OS X	POSIX‡	Matlab
Brainstorm [260]		√√		√√			√√			⇌√						√			
NeuroFEM [261]/Pebbles		√√√		√√√			√√			⇌							√√		
BioPSE [262]/SCIRun [263]		√√√					√√			⇌					√		⇌√		
OpenMEEG		√√		√√			√√			⇌					√		⇌√√√		
Brainvisa/Anatomist [264]										√		√√			√		√√√		
FreeSurfer [265]										√		√√			√		√√		
Caret/Surefit [266]										√		√√			√		√		
Brainsuite [267]										√√√√								√	
EEG/MEG/MRI tlbx* [268]		√		√			√√			⇌⇌⇌⇌√					√		√		
MEG tlbx* [269]				√			√√			√√√√√					√		√		
EEGLAB/FMRILAB [270]							√								√		√		

⇌ stands for Input/Output facility for a feature
†An extensive MR segmentation bibliography is available online [271]
‡POSIX includes all versions of Unix and GNU/Linux
*Matlab Toolbox

Table D.2 Free and Open-source Projects in Python Related to Neuroimaging Data Acquisition and Processing.

Name	Description	URL	PyMVPA
Machine Learning			
Elephant	Multi-purpose library for ML	http://elefant.developer.nicta.com.au	
Shogun	Comprehensive ML toolbox	http://www.shogun-toolbox.org/	✓
Orange	General-purpose data mining	http://www.ailab.si/orange	
PyML	ML in Python	http://pyml.sourceforge.net	
MDP	Modular data processing	http://mdp-toolkit.sourceforge.net	✓
hcluster	Agglomerative clustering	http://code.google.com/p/scipy-cluster/	✓
–	Other Python modules	http://www.mloss.org/software/language/python	
Neuroscience Related			
NiPy	Neuroimaging data analysis	http://neuroimaging.scipy.org	
PyMGH	Access FreeSurfer's *.mgh* files	http://code.google.com/p/pyfsio	
PyNIfTI	Access NIfTI/Analyze files	http://niftilib.sourceforge.net/pynifti	✓
OpenMEEG	EEG/MEG inverse problems	http://www-sop.inria.fr/odyssee/software/OpenMEEG	
Brian	Spiking NN simulation	http://brian.di.ens.fr/	
PyNN	Specification of NN models	http://neuralensemble.org/trac/PyNN	
Stimuli and Experiment Design			
PyEPL	Create complete experiments	http://pyepl.sourceforge.net/	
VisionEgg	Visual Stimuli Generation	http://www.visionegg.org	
PsychoPy	Create psychophysical stimuli	http://www.psychopy.org/	
PIL	Python Imaging Library	http://www.pythonware.com/products/pil/	
Interfaces to Other Computing Environments			
RPy	Interface to R	http://rpy.sourceforge.net/	✓
mlabwrap	Interface to Matlab	http://mlabwrap.sourceforge.net/	
Generic			
Matplotlib	2D Plotting	http://matplotlib.sourceforge.net	✓
Mayavi2	Interactive 3D visualization	http://code.enthought.com/projects/mayavi	
PyExcelerator	Access MS Excel files	http://sourceforge.net/projects/pyexcelerator	
pywavelets	Discrete wavelet transforms	http://www.pybytes.com/pywavelets/	✓

REFERENCES

[1] P. Churchland and T. Sejnowski. *The Computational Brain*. MIT Press, Cambridge, MA, 1992. **xx**, 6

[2] H. Markram. The blue brain project, Feb. 2006. ISSN 1471-003X. 7

[3] D. Cohen. Magnetoencephalography: detection of the brain's electrical activity with a superconducting magnetometer. *Science*, 175:664-6, 1972. 9

[4] Y. O. Halchenko, S. J. Hanson, and B. A. Pearlmutter. *Multimodal Integration: fMRI, MRI, EEG, MEG*, chapter 8, pages 223-265. Signal Processing and Communications. Dekker, 2005. ISBN 0824725425. 9, 88

[5] J. Bodurka and P. A. Bandettini. Toward direct mapping of neuronal activity: MRI detection of ultraweak, transient magnetic field changes. *Magn Reson Med*, 47(6):1052-1058, 2002. 11

[6] J. Xiong, P. T. Fox, and J. H. Gao. Directly mapping magnetic field effects of neuronal activity by magnetic resonance imaging. *Hum Brain Mapp*, 20(1): 41-49, 2003. 11

[7] C. S. Roy and C. S. Sherrington. On the regulation of the blood-supply of the brain. *J Physiol (London)*, 11:85-108, 1890. 11

[8] S. Ogawa, T. Lee, A. Nayak, and P. Glynn. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn Reson Med*, 14(1):68-78, 1990. 11

[9] S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A*, 87 (24):9868-9872, 1990. 11

[10] J. Belliveau, D. Kennedy, Jr, R. McKinstry, B. Buchbinder, R. Weisskoff, M. Cohen, J. Vevea, T. Brady, and B. Rosen. Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, 254(5032):716-719, 1991. 11

[11] B. Rosen, J. Belliveau, H. Aronen, D. Kennedy, B. Buchbinder, A. Fischman, M. Gruber, J. Glas, R. Weisskoff, M. Cohen, and et al. Susceptibility contrast imaging of cerebral blood volume: human experience. *Magn Reson Med*, 22 (2):293-9; discussion 300-3, 1991. 11

[12] R. S. J. Frackowiak, K. J. Friston, R. Dolan, and C. Price. *Human Brain Function*. Academic Press, second edition, November 2003. ISBN 0122648412. 12, 44, 47

[13] A. J. O'Toole, F. Jiang, H. Abdi, N. Penard, J. P. Dunlop, and M. A. Parent. Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *Journal of Cognitive Neuroscience*, 19:1735-1752, 2007. 12, 17, 51, 133

[14] N. Kriegeskorte and P. Bandettini. Analyzing for information, not activation, to exploit high-resolution fMRI. *NeuroImage*, 38(4):649-62, Dec. 2007. ISSN 1053-8119. 12

[15] F. Pereira, T. Mitchell, and M. Botvinick. Machine learning classifiers and fmri: A tutorial overview. *NeuroImage*, in press. 14

[16] S. Sonnenburg, M. Braun, C. S. Ong, S. Bengio, L. Bottou, G. Holmes, Y. LeCun, K.-R. Müller, F. Pereira, C. E. Rasmussen, G. Rätsch, B. Schölkopf, A. Smola, P. Vincent, J. Weston, and R. Williamson. The need for open source software in machine learning. *Journal of Machine Learning Research*, 8:2443-2466, 2007. 14, 126

[17] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, 1995. ISBN 0-387-94559-8. **xx**, 14, 24, 38, 127, 132

[18] Y. Kamitani and F. Tong. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8:679-685, 2005. ISSN 1097-6256. 14, 15

[19] J.-D. Haynes, K. Sakai, G. Rees, S. Gilbert, C. Frith, and R. E. Passingham. Reading hidden intentions in the human brain. *Current Biology*, 17:323-328, 2007. ISSN 0960-9822. 14, 136

[20] N. Kriegeskorte, R. Goebel, and P. Bandettini. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the USA*, 103: 3863-3868, 2006. ISSN 0027-8424. 14, 26, 136, 138

[21] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Science*, 10: 424-430, 2006. ISSN 1364-6613. 14, 142

[22] M. Hanke, Y. O. Halchenko, P. B. Sederberg, S. J. Hanson, J. V. Haxby, and S. Pollmann. PyMVPA: A Python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*, 7(1):37-53, Mar. 2009. 14, 149

[23] M. Hanke, Y. O. Halchenko, P. B. Sederberg, E. Olivetti, I. Fründ, J. W. Rieger, C. S. Herrmann, J. V. Haxby, S. J. Hanson, and S. Pollmann. PyMVPA: A unifying approach to the analysis of neuroscientific data. *Frontiers in Neuroinformatics*, 3(3), 2009. 14, 21, 133, 149

[24] J.-D. Haynes and G. Rees. Predicting the orientation of invisible stimuli from activity in human primary cortex. *Nature Neuroscience*, 8:686-691, 2005. 15

[25] J. Haxby, M. Gobbini, M. Furey, A. Ishai, J. Schouten, and P. Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293:2425-2430, 2001. **xxii**, 15, 48, 49, 50, 51, 52, 53, 57, 60, 66, 127, 133

[26] S. Hanson, T. Matsuka, and J. Haxby. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? *NeuroImage*, 23:156-166, 2004. 15, 29, 48, 50, 51, 58, 59, 62, 133

[27] A. J. O'Toole, F. Jiang, H. Abdi, and J. V. Haxby. Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, 17:580-590, 2005. 15

[28] S. J. Hanson and Y. O. Halchenko. Brain reading using full brain support vector machines for object recognition: there is no "face" identification area. *Neural Computation*, 20:486-503, 2008. ISSN 0899-7667. 15, 51, 133, 149

[29] M. Hanke, Y. O. Halchenko, P. B. Sederberg, and J. M. Hughes. *The PyMVPA Manual*. Available online at http://www.pymvpa.org/PyMVPA-Manual.pdf. 17, 126, 133, 144

[30] G. Detre, S. M. Polyn, C. Moore, V. Natu, B. Singer, J. Cohen, J. V. Haxby, and K. A. Norman. The multi-voxel pattern analysis (MVPA) toolbox. Poster presented at the Annual Meeting of the Organization for Human Brain Mapping (Florence, Italy), 2006. 20

[31] S. LaConte, S. Strother, V. Cherkassky, J. Anderson, and X. Hu. Support vector machines for temporal classification of block design fMRI data. *NeuroImage*, 26:317-329, 2005. ISSN 1053-8119. 20, 51, 62, 142

[32] C. F. Beckmann and S. M. Smith. Tensorial extensions of independent component analysis for multisubject FMRI analysis. *NeuroImage*, 25(1):294-311, Mar. 2005. ISSN 1053-8119. 20

[33] S. Makeig, S. Debener, J. Onton, and A. Delorme. Mining event-related brain dynamics. *Trends Cogn Sci*, 8(5):204-10, May 2004. ISSN 1364-6613. 20

[34] Y. Sun. Iterative RELIEF for feature weighting: Algorithms, theories and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1035-1051, June 2007. **xx**, 24, 38, 127

[35] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46:389-422, 2002. ISSN 0885-6125. **xx**, 24, 30, 127, 139, 140, 141

[36] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning*, 3:1157-1182, 2003. **xx**, 24, 30, 127, 140

[37] B. Krishnapuram, L. Carin, M. A. Figueiredo, and A. J. Hartemink. Sparse multinomial logistic regression: fast algorithms and generalization bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:957-968, 2005. **xx**, 24, 33, 38, 127, 131

[38] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, 1993. **xx**, 24, 27, 129

[39] T. E. Nichols and A. P. Holmes. Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15:1-25, 2001. 27

[40] X. Chen, F. Pereira, W. Lee, S. Strother, and T. Mitchell. Exploring predictive and reproducible modeling with the single-subject FIAC dataset. *Human Brain Mapping*, 27:452-461, 2006. 29

[41] A. Rakotomamonjy. Variable selection using SVM-based criteria. *Journal of Machine Learning Research*, 3:1357-1370, 2003. 29, 30, 63

[42] S. L. Sally and J. B. Kelly. Organization of auditory cortex in the albino rat: sound frequency. *J Neurophysiol*, 59(5):1627-38, May 1988. ISSN 0022-3077. 32

[43] T. C. Handy, editor. MIT Press, Cambridge, MA, Oct. 2004. ISBN 0262083337. 34

[44] T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski. Analyzing and visualizing single-trial event-related potentials. In *Adv in Neu Info Proc Sys 11*, pages 118-24. MIT Press, 1999. 37

[45] A. C. Tang, B. A. Pearlmutter, D. B. Phung, and N. A. Malaszenko. Independent components of magnetoencephalography: Single-trial response onset times. *NeuroImage*, 2002. 37

[46] B. A. Pearlmutter, A. C. Tang, M. Zibulevsky, T. A. Hely, and M. P. Weisend. An MEG study of response latency and variability in the human visual system during a simple visual-motor integration task. *Soc for Neurosci Abstracts*, 29 (766.8), 1999. 37

[47] B. A. Pearlmutter, S. Jaramillo, A. C. Tang, and G. Nolte. Relationship between the phase of alpha-band oscillation and single-trial visually evoked responses. *Soc for Neurosci Abstracts*, 31(821.1), 2001. 37

[48] L. Parra, C. Alvino, A. Tang, B. Pearlmutter, N. Yeung, A. Osman, and P. Sajda. Linear spatial integration for single-trial detection in encephalography. *NeuroImage*, 17(1):223-30, Sep. 2002. 37

[49] I. Fründ, N. A. Busch, J. Schadow, T. Gruber, U. Körner, and C. S. Herrmann. Time pressure modulates electrophysiological correlates of early visual processing. *PLoS ONE*, 3(2):e1675, 2008. ISSN 1932-6203. 37, 38, 39

[50] N. A. Busch, C. S. Herrmann, M. M. Müller, D. Lenz, and T. Gruber. A cross-laboratory study of event-related gamma activity in a standard object recognition paradigm. *NeuroImage*, 33:1169-1177, 2006. 38

[51] C. S. Herrmann, D. Lenz, S. Junge, N. A. Busch, and B. Maess. Memory-matches evoke human gamma-responses. *BMC Neurosci*, 5(13), 2004. 38

[52] C. E. Rasmussen and C. K. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 2006. 38

[53] J.-P. Lachaux, N. George, C. Tallon-Baudry, J. Martinerie, L. Hugueville, L. Minotti, P. Kahane, and B. Renault. The many faces of the gamma band response to complex visual stimuli. *NeuroImage*, 25:491-501, 2005. 39

[54] J. W. Rieger, C. Reichert, K. R. Gegenfurtner, T. Noesselt, C. Braun, H.-J. Heinze, R. Kruse, and H. Hinrichs. Predicting the recognition of natural scenes from single trial MEG recordings of brain activity. *NeuroImage*, 42(3):1056-68, Sep. 2008. ISSN 1095-9572. 41, 42, 43

[55] J. W. Rieger, C. Braun, H. H. Bülthoff, and K. R. Gegenfurtner. The dynamics of visual pattern masking in natural scene processing: A magnetoencephalography study. *J. Vis.*, 5(3):275-286, 3 2005. ISSN 1534-7362. 41

[56] K. Veropoulos, C. Campbell, and N. Cristianini. Controlling the sensitivity of support vector machines. In *Proceedings of the International Joint Conference on AI*, pages 55-60, 1999. 42

[57] N. Kanwisher, J. McDermott, and M. M. Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*, 17 (11):4302-11, 1997. 46, 48, 62

[58] N. Kanwisher, D. Stanley, and A. Harris. The fusiform face area is selective for faces not animals. *Neuroreport*, 10(1):183-7, 1999. 46

[59] M. Spiridon and N. Kanwisher. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron*, 35(6):1157-65, 2002. 46, 48

[60] N. Kanwisher. Neuroscience. What's in a face? *Science*, 311(5761):617-8, 2006. 48, 69

[61] I. Gauthier, P. Skudlarski, J. C. Gore, and A. W. Anderson. Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci*, 3(2): 191-7, 2000. 48

[62] D. H. Krantz, R. D. Luce, P. Suppes, and A. Tversky. *Foundations of Measurement*, volume 1. Academic Press, San Diego, 1971. 49

[63] R. D. Luce. *Individual choice behavior: a theoretical analysis*. Wiley, N.Y., 1959. 49

[64] K. Grill-Spector and N. Kanwisher. Visual recognition: as soon as you know it is there, you know what it is. *Psychol Sci*, 16(2):152-60, 2005. 50

[65] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Mach Learn*, 46(1-3):389-422, 2002. ISSN 0885-6125. 50, 58, 63

[66] S. J. Hanson and B. M. Bly. The distribution of BOLD susceptibility effects in the brain is non-Gaussian. *Neuroreport*, 12(9):1971-7, 2001. 50

[67] C. C. Chen, C. W. Tyler, and H. A. Baseler. Statistical properties of BOLD magnetic resonance activity in the human brain. *NeuroImage*, 20(2):1096-109, 2003. 50

[68] A. M. Wink and J. B. T. M. Roerdink. BOLD noise assumptions in fMRI. *Int. J. Biomed. Imaging*, 2006, 2006. 50

[69] S. J. Hanson and D. J. Burr. What connectionist models learn: Toward a theory of representation in connectionist networks. *Behavioral and Brain Sciences*, 13: 471-518, 1990. 51

[70] T. Mitchell, R. Hutchinson, R. S. Niculescu, F. Pereira, X. Wang, M. Just, and S. Newman. Learning to decode cognitive states from brain images. *Machine Learning*, 57:145-175, 2004. 51

[71] L. L. Chao, A. Martin, and J. V. Haxby. Are face-responsive regions selective only for faces? *Neuroreport*, 10(14):2945-50, 1999. 51

[72] I. Gauthier, M. J. Tarr, J. Moylan, P. Skudlarski, J. C. Gore, and A. W. Anderson. The fusiform "face area" is part of a network that processes faces at the individual level. *J Cogn Neurosci*, 12(3):495-504, 2000. 51

[73] R. Malach, I. Levy, and U. Hasson. The topography of high-order human object areas. *Trends Cogn Sci*, 6(4):176-184, 2002. 51

[74] M. Jenkinson, P. Bannister, J. Brady, and S. Smith. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17:825-841, 2002. 53

[75] D. Flitney, M. Webster, B. Patenaude, L. Seidman, J. Goldstein, D. Tordesillas Gutierrez, S. Eickhoff, K. Amunts, K. Zilles, J. Lancaster, C. Haselgrove, D. Kennedy, M. Jenkinson, and S. Smith. Anatomical brain atlases and their application in the FSLView visualisation tool. In *Thirteenth Annual Meeting of the Organization for Human Brain Mapping*, 2007. 53

[76] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. M. Brady, and P. M. Matthews. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23 Suppl 1:S208-19, 2004. ISSN 1053-8119. 53, 60

[77] I. Guyon, B. Boser, and V. Vapnik. Automatic capacity tuning of very large VC-dimension classifiers. In S. J. Hanson, J. D. Cowan, and C. L. Giles, editors, *Advances in Neural Information Processing Systems*, volume 5, pages 147-155. Morgan Kaufmann, San Mateo, CA, 1993. 56

[78] K. J. Friston, P. Rotshtein, J. J. Geng, P. Sterzer, and R. N. Henson. A critique of functional localisers. *NeuroImage*, 30(4):1077-87, 2006. 62

[79] R. Saxe, M. Brett, and N. Kanwisher. Divide and conquer: a defense of functional localizers. *NeuroImage*, 30(4):1088-96; discussion 1097-9, 2006. 62

[80] B. Ishak and B. Ghattas. An efficient method for variable selection using SVM based criteria. *Journal of Machine Learning Research*, 6, 2005. 63

[81] R. J. Hyndman and Y. Fan. Sample quantiles in statistical packages. *The American Statistician*, 50(4):361-366, 1996. 63

[82] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini. The distributed human neural system for face perception. *Trends Cogn Sci*, 4(6):223-233, 2000. 68

[83] F. Lazeyras, I. Zimine, O. Blanke, S. H. Perrig, and M. Seeck. Functional MRI with simultaneous EEG recording: feasibility and application to motor and visual activation. *J Magn Reson Imaging*, 13(6):943-948, 2001. 71, 76

[84] K. D. Singh, G. R. Barnes, A. Hillebrand, E. M. Forde, and A. L. Williams. Task-related changes in cortical synchronization are spatially coincident with the hemodynamic response. *NeuroImage*, 16(1):103-114, 2002. 71

[85] M. Sommer, J. Meinhardt, and H. P. Volz. Combined measurement of event-related potentials (ERPs) and fMRI. *Acta Neurobiol Exp (Wars)*, 63(1):49-53, 2003. 71, 77

[86] S. Vanni, J. Warnking, M. Dojat, C. Delon-Martin, J. Bullier, and C. Segebarth. Sequence of pattern onset responses in the human visual areas: an fMRI constrained VEP source analysis. *NeuroImage*, 21(3):801-817, 2004. 71

[87] K. Whittingstall, G. Stroink, and M. Schmidt. Evaluating the spatial relationship of event-related potential and functional MRI sources in the primary visual cortex. *Hum Brain Mapp*, 2006. 71

[88] Y. Nagai, H. D. Critchley, E. Featherstone, P. B. Fenwick, M. R. Trimble, and R. J. Dolan. Brain activity relating to the contingent negative variation: an fMRI investigation. *NeuroImage*, 21(4):1232-41, 2004. 71, 86

[89] A. Korvenoja, J. Huttunen, E. Salli, H. Pohjonen, S. Martinkauppi, J. M. Palva, L. Lauronen, J. Virtanen, R. J. Ilmoniemi, and H. J. Aronen. Activation of multiple cortical areas in response to somatosensory stimulation: combined magnetoencephalographic and functional magnetic resonance imaging. *Hum Brain Mapp*, 8(1):13-27, 1999. 71, 88, 89

[90] M. Schulz, W. Chau, S. J. Graham, A. R. McIntosh, B. Ross, R. Ishii, and C. Pantev. An integrative MEG-fMRI study of the primary somatosensory cortex using cross-modal correspondence analysis. *NeuroImage*, 22(1):120-133, 2004. 71, 76, 89

[91] M. S. Cohen, R. I. Goldman, J. Stern, and J. Engel, Jr. Simultaneous EEG and fMRI made easy. *NeuroImage*, 13(6 Supp.1):S6, Jan. 2001. 71, 73, 74

[92] R. I. Goldman, J. M. Stern, J. Engel, Jr, and M. S. Cohen. Simultaneous EEG and fMRI of the alpha rhythm. *Neuroreport*, 13(18):2487-2492, 2002. 71, 86, 88, 122

[93] M. Moosmann, P. Ritter, I. Krastel, A. Brink, S. Thees, F. Blankenburg, B. Taskin, H. Obrig, and A. Villringer. Correlates of alpha rhythm in functional magnetic resonance imaging and near infrared spectroscopy. *NeuroImage*, 20(1):145-158, 2003. 71

[94] H. Laufs, A. Kleinschmidt, A. Beyerle, E. Eger, A. Salek-Haddadi, C. Preibisch, and K. Krakow. EEG-correlated fMRI of human alpha activity. *NeuroImage*, 19(4):1463-1476, 2003. 71, 86

[95] M. J. Makiranta, J. Ruohonen, K. Suominen, E. Sonkajarvi, T. Salomaki, V. Kiviniemi, T. Seppanen, S. Alahuhta, V. Jantti, and O. Tervonen. BOLD-contrast functional MRI signal changes related to intermittent rhythmic delta activity in EEG during voluntary hyperventilation-simultaneous EEG and fMRI study. *NeuroImage*, 22(1):222-231, 2004. 71, 77

[96] J. Foucher, H. Otzenberger, and D. Gounot. Where arousal meets attention: a simultaneous fMRI and EEG recording study. *NeuroImage*, 22(2):688-697, 2004. 71

[97] S. G. Horovitz, P. Skudlarski, and J. C. Gore. Correlations and dissociations between BOLD signal and P300 amplitude in an auditory oddball task: a parametric approach to combining fMRI and ERP. *Magn Reson Imaging*, 20 (4):319-325, 2002. 71, 76, 85

[98] M. Brazdil, M. Dobsik, M. Mikl, P. Hlustik, P. Daniel, M. Pazourkova, P. Krupa, and I. Rektor. Combined event-related fMRI and intracerebral ERP study of an auditory oddball task. *NeuroImage*, 26(1):285-93, 2005. 71

[99] T. Eichele, K. Specht, M. Moosmann, M. L. Jongsma, R. Q. Quiroga, H. Nordby, and K. Hugdahl. Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. *Proc Natl Acad Sci U S A*, 102(49):17798-803, 2005. 71

[100] E. Liebenthal, M. L. Ellingson, M. V. Spanaki, T. E. Prieto, K. M. Ropella, and J. R. Binder. Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. *NeuroImage*, 19(4):1395-1404, 2003. 71, 86

[101] 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(6):1327-1336, 2001. 71

[102] P. Sehatpour, S. Molholm, D. Javitt, and J. Foxe. Spatiotemporal dynamics of human object recognition processing: An integrated high-density electrical mapping and functional imaging study of ”closure” processes. *NeuroImage*, 2005. 71

[103] V. Menon, J. M. Ford, K. O. Lim, G. H. Glover, and A. Pfefferbaum. Combined event-related fMRI and EEG evidence for temporal-parietal cortex activation during target detection. *Neuroreport*, 8(14):3029-3037, 1997. 71, 76

[104] C. Mulert, L. Jager, R. Schmitt, P. Bussfeld, O. Pogarell, H. J. Moller, G. Juckel, and U. Hegerl. Integration of fMRI and simultaneous EEG: towards a comprehensive understanding of localization and time-course of brain activity in target detection. *NeuroImage*, 22(1):83-94, 2004. 71

[105] S. G. Horovitz, B. Rossion, P. Skudlarski, and J. C. Gore. Parametric design and correlational analyses help integrating fMRI and electrophysiological data during face processing. *NeuroImage*, 22(4):1587-1595, 2004. 71, 76, 85, 86

[106] F. R. Huang-Hellinger, H. C. Breiter, G. McCormack, M. S. Cohen, K. K. Kwong, J. P. Sutton, R. L. Savoy, R. M. Weisskoff, T. L. Davis, J. R. Baker, J. W. Belliveau, and B. R. Rosen. Simultaneous functional magnetic resonance imaging and electrophysiological recording. *Hum Brain Mapp*, 3:13-25, 1995. 71, 77

[107] D. Vitacco, D. Brandeis, R. Pascual-Marqui, and E. Martin. Correspondence of event-related potential tomography and functional magnetic resonance imaging during language processing. *Hum Brain Mapp*, 17(1):4-12, 2002. 71, 85

[108] P. Grummich, C. Nimsky, E. Pauli, M. Buchfelder, and O. Ganslandt. Combining fMRI and MEG increases the reliability of presurgical language localization: A clinical study on the difference between and congruence of both modalities. *NeuroImage*, 2006. 71

[109] T. Hinterberger, N. Weiskopf, R. Veit, B. Wilhelm, E. Betta, and N. Birbaumer. An EEG-driven brain-computer interface combined with functional magnetic resonance imaging (fMRI). *IEEE Trans Biomed Eng*, 51(6):971-4, 2004. 71

[110] S. Warach, J. R. Ives, G. Schlaug, M. R. Patel, D. G. Darby, V. Thangaraj, R. R. Edelman, and D. L. Schomer. EEG-triggered echo-planar functional MRI in epilepsy. *Neurology*, 47(1):89-93, 1996. 71, 77

[111] M. Seeck, F. Lazeyras, C. M. Michel, O. Blanke, C. A. Gericke, J. Ives, J. Delavelle, X. Golay, C. A. Haenggeli, N. de Tribolet, and T. Landis. Non-invasive epileptic focus localization using EEG-triggered functional MRI and electromagnetic tomography. *Electroencephalogr Clin Neurophysiol*, 106(6): 508-512, 1998. 71, 77

[112] K. Krakow, U. C. Wieshmann, F. G. Woermann, M. R. Symms, M. A. McLean, L. Lemieux, P. J. Allen, G. J. Barker, D. R. Fish, and J. S. Duncan. Multimodal MR imaging: functional, diffusion tensor, and chemical shift imaging in a patient with localization-related epilepsy. *Epilepsia*, 40(10):1459-1462, 1999. 71

[113] K. Krakow, F. G. Woermann, M. R. Symms, P. J. Allen, L. Lemieux, G. J. Barker, J. S. Duncan, and D. R. Fish. EEG-triggered functional MRI of interictal epileptiform activity in patients with partial seizures. *Brain*, 122(9):1679-1688, 1999. 71

[114] K. Krakow, L. Lemieux, D. Messina, C. A. Scott, M. R. Symms, J. S. Duncan, and D. R. Fish. Spatio-temporal imaging of focal interictal epileptiform activity using EEG-triggered functional MRI. *Epileptic Disord*, 3(2):67-74, 2001. 71, 77

[115] G. Lantz, L. Spinelli, R. G. Menendez, M. Seeck, and C. M. Michel. Localization of distributed sources and comparison with functional MRI. *Epileptic Disord*, Special Issue:45-58, 2001. 71

[116] L. Lemieux, K. Krakow, and D. R. Fish. Comparison of spike-triggered functional MRI BOLD activation and EEG dipole model localization. *NeuroImage*, 14 (5):1097-1104, 2001. 71, 88

[117] A. Waites, M. Shaw, R. Briellmann, A. Labate, D. Abbott, and G. Jackson. How reliable are fMRI-EEG studies of epilepsy? A nonparametric approach to analysis validation and optimization. *NeuroImage*, 24(1):192-199, 2005. 71

[118] A. D. Liston, J. C. De Munck, K. Hamandi, H. Laufs, P. Ossenblok, J. S. Duncan, and L. Lemieux. Analysis of EEG-fMRI data in focal epilepsy based on automated spike classification and Signal Space Projection. *NeuroImage*, 2006. 71

[119] Y. Lu, C. Grova, E. Kobayashi, F. Dubeau, and J. Gotman. Using voxel-specific hemodynamic response function in EEG-fMRI data analysis: An estimation and detection model. *NeuroImage*, 34(1):195-203, 2007. 71, 80

[120] S. M. Mirsattari, Z. Wang, J. R. Ives, F. Bihari, L. S. Leung, R. Bartha, and R. S. Menon. Linear aspects of transformation from interictal epileptic discharges to BOLD fMRI signals in an animal model of occipital epilepsy. *NeuroImage*, 30(4):1133-1148, 2006. 71

[121] P. Ritter and A. Villringer. Simultaneous EEG-fMRI. *Neurosci Biobehav Rev*, 30 (6):823-38, 2006. 72, 74

[122] G. Sammer, C. Blecker, H. Gebhardt, P. Kirsch, R. Stark, and D. Vaitl. Acquisition of typical EEG waveforms during fMRI: SSVEP, LRP, and frontal theta. *NeuroImage*, 24(4):1012-1024, 2005. 72

[123] L. K. Hansen, F. A. Nielsen, S. C. Strother, and N. Lange. Consensus inference in neuroimaging. *NeuroImage*, 13(6 Pt 1):1212-8, 2001. 72

[124] C. H. Im, Z. Liu, N. Zhang, W. Chen, and B. He. Functional cortical source imaging from simultaneously recorded ERP and fMRI. *J Neurosci Methods*, 2006. 72

[125] H. Laufs, J. Daunizeau, D. W. Carmichael, and A. Kleinschmidt. Recent advances in recording electrophysiological data simultaneously with magnetic resonance imaging. *NeuroImage*, 2007. 73

[126] J. R. Ives, S. Warach, F. Schmitt, R. R. Edelman, and D. L. Schomer. Monitoring the patient's EEG during echo planar MRI. *Electroencephalogr Clin Neurophysiol*, 87(6):417-420, 1993. 73, 74, 76

[127] M. C. Schmid, A. Oeltermann, C. Juchem, N. K. Logothetis, and S. M. Smirnakis. Simultaneous EEG and fMRI in the macaque monkey at 4.7 Tesla. *Magn Reson Imaging*, 24(4):335-42, 2006. 73

[128] P. J. Allen, O. Josephs, and R. Turner. A method for removing imaging artifact from continuous EEG recorded during functional MRI. *NeuroImage*, 12(2): 230-239, 2000. 73

[129] A. Salek-Haddadi, M. Merschhemke, L. Lemieux, and D. R. Fish. Simultaneous EEG-Correlated Ictal fMRI. *NeuroImage*, 16(1):32-40, 2002. 73, 77

[130] M. S. Cohen. Method and apparatus for reducing contamination of an electrical signal. United States Patent application 0040097802, 2004. 73

[131] K. Anami, T. Mori, F. Tanaka, Y. Kawagoe, J. Okamoto, M. Yarita, T. Ohnishi, M. Yumoto, H. Matsuda, and O. Saitoh. Stepping stone sampling for retrieving artifact-free electroencephalogram during functional magnetic resonance imaging. *NeuroImage*, 19(2.1):281-295, 2003. 73

[132] H. Mandelkow, P. Halder, P. Boesiger, and D. Brandeis. Synchronization facilitates removal of MRI artefacts from concurrent EEG recordings and increases usable bandwidth. *NeuroImage*, 32(3):1120-6, 2006. 74

[133] S. I. Goncalves, P. J. Pouwels, J. P. Kuijer, R. M. Heethaar, and J. C. de Munck. Artifact removal in co-registered EEG/fMRI by selective average subtraction. *Clin Neurophysiol*, 2007. 74

[134] R. I. Goldman, J. M. Stern, J. Engel, Jr, and M. S. Cohen. Acquiring simultaneous EEG and functional MRI. *Clin Neurophysiol*, 111(11):1974-1980, 2000. 74, 76, 77

[135] M. Negishi, M. Abildgaard, T. Nixon, and R. Todd Constable. Removal of time-varying gradient artifacts from EEG data acquired during continuous fMRI. *Clin Neurophysiol*, 115(9):2181-2192, 2004. 74

[136] R. K. Niazy, C. F. Beckmann, G. D. Iannetti, J. M. Brady, and S. M. Smith. Removal of FMRI environment artifacts from EEG data using optimal basis sets. *NeuroImage*, 28(3):720-37, 2005. 74

[137] H. Otzenberger, D. Gounot, and J. R. Foucher. Optimisation of a post-processing method to remove the pulse artifact from EEG data recorded during fMRI: an application to P300 recordings during e-fMRI. *Neurosci Res*, 57(2):230-9, 2007. 74

[138] J. Sijbers, I. Michiels, M. Verhoye, J. Van Audekerke, A. Van der Linden, and D. Van Dyck. Restoration of MR-induced artifacts in simultaneously recorded MR/EEG data. *Magn Reson Imaging*, 17(9):1383-1391, 1999. 74

[139] G. Bonmassar, P. L. Purdon, I. P. Jaaskelainen, K. Chiappa, V. Solo, E. N. Brown, and J. W. Belliveau. Motion and ballistocardiogram artifact removal for interleaved recording of EEG and EPs during MRI. *NeuroImage*, 16(4): 1127-1141, 2002. 74

[140] G. Garreffa, M. Carni, G. Gualniera, G. B. Ricci, L. Bozzao, D. De Carli, P. Morasso, P. Pantano, C. Colonnese, V. Roma, and B. Maraviglia. Real-time MR artifacts filtering during continuous EEG/fMRI acquisition. *Magn Reson Imaging*, 21(10):1175-1189, 2003. 74

[141] G. Srivastava, S. Crottaz-Herbette, K. Lau, G. Glover, and V. Menon. ICA-based procedures for removing ballistocardiogram artifacts from EEG data acquired in the MRI scanner. *NeuroImage*, 24(1):50-60, 2005. 74

[142] K. H. Kim, H. W. Yoon, and H. W. Park. Improved ballistocardiac artifact removal from the electroencephalogram recorded in fMRI. *J Neurosci Methods*, 135 (1-2):193-203, 2004. 74

[143] R. A. Hill, K. H. Chiappa, F. Huang-Hellinger, and B. G. Jenkins. EEG during MR imaging: differentiation of movement artifact from paroxysmal cortical activity. *Neurology*, 45(10):1942-1943, 1995. 74

[144] 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. *Magn Reson Med*, 44(2):277-282, 2000. 74

[145] S. Debener, A. Strobel, B. Sorger, J. Peters, C. Kranczioch, A. K. Engel, and R. Goebel. Improved quality of auditory event-related potentials recorded simultaneously with 3-T fMRI: Removal of the ballistocardiogram artefact. *NeuroImage*, 34(2):587-97, 2007. 74

[146] M. Siniatchkin, F. Moeller, J. Jacobs, U. Stephani, R. Boor, S. Wolff, O. Jansen, H. Siebner, and M. Scherg. Spatial filters and automated spike detection based on brain topographies improve sensitivity of EEG-fMRI studies in focal epilepsy. *NeuroImage*, 37(3):834-43, 2007. 74

[147] L. Lemieux, P. J. Allen, F. Franconi, M. R. Symms, and D. R. Fish. Recording of EEG during fMRI experiments: patient safety. *Magn Reson Med*, 38(6): 943-952, 1997. 74

[148] K. Krakow, P. J. Allen, M. R. Symms, L. Lemieux, O. Josephs, and D. R. Fish. EEG recording during fMRI experiments: image quality. *Hum Brain Mapp*, 10(1): 10-15, 2000. 74, 76

[149] L. M. Angelone, A. Potthast, F. Segonne, S. Iwaki, J. W. Belliveau, and G. Bonmassar. Metallic electrodes and leads in simultaneous EEG-MRI: specific absorption rate (SAR) simulation studies. *Bioelectromagnetics*, 25 (4):285-295, 2004. 76

[150] G. Bonmassar, D. P. Schwartz, A. K. Liu, K. K. Kwong, A. M. Dale, and J. W. Belliveau. Spatiotemporal brain imaging of visual-evoked activity using interleaved EEG and fMRI recordings. *NeuroImage*, 13(6.1):1035-1043, 2001. 76, 77

[151] F. Lazeyras, O. Blanke, S. Perrig, I. Zimine, X. Golay, J. Delavelle, C. M. Michel, N. de Tribolet, J. G. Villemure, and M. Seeck. EEG-triggered functional MRI in patients with pharmacoresistant epilepsy. *J Magn Reson Imaging*, 12(1): 177-185, 2000. 77

[152] D. L. Schomer, G. Bonmassar, F. Lazeyras, M. Seeck, A. Blum, K. Anami, D. Schwartz, J. W. Belliveau, and J. Ives. EEG-Linked functional magnetic resonance imaging in epilepsy and cognitive neurophysiology. *J Clin Neurophysiol*, 17(1):43-58, 2000. 78

[153] S. Ogawa, R. S. Menon, D. W. Tank, S. . G. Kim, H. Merkle, J. M. Ellermann, and K. Ugurbil. Functional brain mapping by blood oxygenation level-dependent

contrast magnetic resonance imaging. a comparison of signal characteristics with a biophysical model. *Biophysical Journal*, 1993. 78

[154] R. B. Buxton and L. R. Frank. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J Cereb Blood Flow Metab*, 17(1):64-72, 1997. 78, 82

[155] T. Obata, T. Liu, K. Miller, W. Luh, E. Wong, L. Frank, and R. Buxton. Discrepancies between BOLD and flow dynamics in primary and supplementary motor areas: application of the balloon model to the interpretation of BOLD transients. *NeuroImage*, 21(1):144-53, 2004. 78

[156] K. E. Stephan, N. Weiskopf, P. M. Drysdale, P. A. Robinson, and K. J. Friston. Comparing hemodynamic models with DCM. *NeuroImage*, 2007. 78

[157] A. Seiyama, J. Seki, H. C. Tanabe, I. Sase, A. Takatsuki, S. Miyauchi, H. Eda, S. Hayashi, T. Imaruoka, T. Iwakura, and T. Yanagida. Circulatory basis of fMRI signals: relationship between changes in the hemodynamic parameters and BOLD signal intensity. *NeuroImage*, 21(4):1204-1214, 2004. 78, 83, 84

[158] T. Deneux and O. Faugeras. Using nonlinear models in fMRI data analysis: Model selection and activation detection. *NeuroImage*, 32(4):1669-89, 2006. 78, 80

[159] J. Kilner, J. Mattout, R. Henson, and K. Friston. Hemodynamic correlates of EEG: A heuristic. *NeuroImage*, 28(1):280-6, 2005. 78

[160] K. J. Friston, P. Jezzard, and R. Turner. Analysis of functional MRI time-series. *Hum Brain Mapp*, 1:153-171, 1994. 79, 80

[161] G. M. Boynton, S. A. Engel, G. H. Glover, and D. J. Heeger. Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci*, 16(13): 4207-4221, 1996. 79

[162] M. S. Cohen. Parametric analysis of fMRI data using linear systems methods. *NeuroImage*, 6(2):93-103, 1997. 79

[163] N. Lange and S. L. Zeger. Non-linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging. *Appl Stat*, 46(1):1-29, 1997. 80

[164] G. H. Glover. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, 9(4):416-429, 1999. 80, 81

[165] J. C. Rajapakse, F. Kruggel, J. M. Maisog, and D. Y. von Cramon. Modeling hemodynamic response for analysis of functional MRI time-series. *Hum Brain Mapp*, 6(4):283-300, 1998. 80

[166] P. Ciuciu, J. B. Poline, G. Marrelec, J. Idier, C. Pallier, and H. Benali. Unsupervised robust nonparametric estimation of the hemodynamic response function for any fMRI experiment. *IEEE Trans Med Imaging*, 22(10):1235-1251, 2003. 80

[167] D. R. Gitelman, W. D. Penny, J. Ashburner, and K. J. Friston. Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. *NeuroImage*, 19(1):200-207, 2003. 80

[168] G. Marrelec, H. Benali, P. Ciuciu, M. Pelegrini-Issac, and J. B. Poline. Robust Bayesian estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information. *Hum Brain Mapp*, 19(1): 1-17, 2003. 80

[169] Y. Lu, A. P. Bagshaw, C. Grova, E. Kobayashi, F. Dubeau, and J. Gotman. Using voxel-specific hemodynamic response function in EEG-fMRI data analysis. *NeuroImage*, 32(1):238-47, 2006. 80

[170] D. A. Soltysik, K. K. Peck, K. D. White, B. Crosson, and R. W. Briggs. Comparison of hemodynamic response nonlinearity across primary cortical areas. *NeuroImage*, 22(3):1117-1127, 2004. 81

[171] D. Attwell and C. Iadecola. The neural basis of functional brain imaging signals. *Trends Neurosci*, 25(12):621-625, 2002. 81

[172] 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 Topogr*, 13(2):79-96, 2000. 82

[173] N. K. Logothetis and B. A. Wandell. Interpreting the BOLD signal. *Annu Rev Physiol*, 66:735-769, 2004. 82, 97

[174] O. J. Arthurs and S. Boniface. How well do we understand the neural origins of the fMRI BOLD signal? *Trends Neurosci*, 25(1):27-31, 2002. 82

[175] X. Wan, J. Riera, K. Iwata, M. Takahashi, T. Wakabayashi, and R. Kawashima. The neural basis of the hemodynamic response nonlinearity in human primary visual cortex: Implications for neurovascular coupling mechanism. *NeuroImage*, 2006. 82

[176] K. J. Friston. Bayesian estimation of dynamical systems: an application to fMRI. *NeuroImage*, 16(2):513-30, 2002. 82

[177] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843):150-157, 2001. 83

[178] D. S. Kim, I. Ronen, C. Olman, S. G. Kim, K. Ugurbil, and L. J. Toth. Spatial relationship between neuronal activity and BOLD functional MRI. *NeuroImage*, 21(3):876-85, 2004. 83

[179] T. J. Huppert, R. D. Hoge, S. G. Diamond, M. A. Franceschini, and D. A. Boas. A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage*, 29(2):368-82, 2006. 83

[180] A. Devor, A. K. Dunn, M. L. Andermann, I. Ulbert, D. A. Boas, and A. M. Dale. Coupling of total hemoglobin concentration, oxygenation, and neural activity in rat somatosensory cortex. *Neuron*, 39(2):353-359, 2003. 83

[181] R. S. Paulesu, R. S. J. Frackowiak, and G. Bottini. Maps of somatosensory systems. In R. S. J. Frackowiak, editor, *Human brain function*, page 528. Academic Press, San Diego, CA, 1997. 83

[182] J. K. Choi, Y. I. Chen, E. Hamel, and B. G. Jenkins. Brain hemodynamic changes mediated by dopamine receptors: Role of the cerebral microvasculature in dopamine-mediated neurovascular coupling. *NeuroImage*, 2006. 84

[183] B. Stefanovic, J. M. Warnking, and G. B. Pike. Hemodynamic and metabolic responses to neuronal inhibition. *NeuroImage*, 22(2):771-778, 2004. 84

[184] D. Nair. About being BOLD. *Brain Res Brain Res Rev*, 2005. 84

[185] B. Diehl, A. Salek-haddadi, D. R. Fish, and L. Lemieux. Mapping of spikes, slow waves, and motor tasks in a patient with malformation of cortical development using simultaneous EEG and fMRI. *Magn Reson Imaging*, 21(10):1167-73, 2003. 86

[186] E. Niedermeyer. Alpha rhythms as physiological and abnormal phenomena. *Int J Psychophysiol*, 26(1-3):31-49, 1997. 86

[187] S. Goncalves, J. de Munck, P. Pouwels, R. Schoonhoven, J. Kuijer, N. Maurits, J. Hoogduin, E. Van Someren, R. Heethaar, and F. Lopes da Silva. Correlating the alpha rhythm to BOLD using simultaneous EEG/fMRI: Inter-subject variability. *NeuroImage*, 2005. 86, 122

[188] H. Laufs, K. Krakow, P. Sterzer, E. Eger, A. Beyerle, A. Salek-Haddadi, and A. Kleinschmidt. Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proc Natl Acad Sci U S A*, 100(19):11053-11058, 2003. 86

[189] J. C. de Munck, S. I. Goncalves, L. Huijboom, J. P. Kuijer, P. J. Pouwels, R. M. Heethaar, and F. H. Lopes da Silva. The hemodynamic response of the alpha rhythm: an EEG/fMRI study. *NeuroImage*, 35(3):1142-51, 2007. 87

[190] E. Martinez-Montes, P. A. Valdes-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen. Concurrent EEG/fMRI analysis by multiway Partial Least Squares. *NeuroImage*, 22(3):1023-1034, 2004. 87

[191] S. P. Ahlfors, G. V. Simpson, A. M. Dale, J. W. Belliveau, A. K. Liu, A. Korvenoja, J. Virtanen, M. Huotilainen, R. B. Tootell, H. J. Aronen, and R. J. Ilmoniemi. Spatiotemporal activity of a cortical network for processing visual motion revealed by MEG and fMRI. *J Neurophysiol*, 82(5):2545-2555, 1999. 88

[192] R. Beisteiner, M. Erdler, C. Teichtmeister, M. Diemling, E. Moser, V. Edward, and L. Deecke. Magnetoencephalography may help to improve functional MRI brain mapping. *Eur J Neurosci*, 9(5):1072-1077, 1997. 88

[193] S. L. Gonzalez Andino, O. Blanke, G. Lantz, G. Thut, and R. Grave de Peralta Menendez. The use of functional constraints for the neuroelectromagnetic inverse problem: Alternatives and caveats. *Int J Bioelectromag*, 3(1), 2001. 88, 91

[194] C. Christmann, M. Ruf, D. F. Braus, and H. Flor. Simultaneous electroencephalography and functional magnetic resonance imaging of primary and secondary somatosensory cortex in humans after electrical stimulation. *Neurosci Lett*, 333(1):69-73, 2002. 88

[195] H. Kober, C. Nimsky, M. Moller, P. Hastreiter, R. Fahlbusch, and O. Ganslandt. Correlation of sensorimotor activation with functional magnetic resonance imaging and magnetoencephalography in presurgical functional imaging: a spatial analysis. *NeuroImage*, 14(5):1214-1228, 2001. 89

[196] M. Wagner and M. Fuchs. Integration of functional MRI, structural MRI, EEG, and MEG. *Int J Bioelectromag*, 3(1), 2001. 89, 91

[197] K. Mathiak and A. Fallgatter. Combining magnetoencephalography and functional magnetic resonance imaging. *Int Rev Neurobiol*, 68:121-48, 2005. 89

[198] J. J. Foxe, M. E. McCourt, and D. C. Javitt. Right hemisphere control of visuospatial attention: line-bisection judgments evaluated with high-density electrical mapping and source analysis. *NeuroImage*, 19(3):710-726, 2003. 89

[199] A. M. Dale and M. I. Sereno. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *J Cog Neurosci*, 5(2):162-176, 1993. 89

[200] 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(5):406-431, 1995. 90

[201] A. K. Liu, J. W. Belliveau, and A. M. Dale. Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulations. *Proc Natl Acad Sci U S A*, 95(15):8945-8950, 1998. 90

[202] S. P. Ahlfors and G. V. Simpson. Geometrical interpretation of fMRI-guided MEG/EEG inverse estimates. *NeuroImage*, 22(1):323-332, 2004. **xxiv**, 90, 92

[203] A. M. Dale, A. K. Liu, B. Fischl, J. D. Lewine, R. L. Buckner, J. W. Belliveau, and E. Halgren. Dynamic statistical parameter mapping: combining fMRI and MEG to produce high resolution imaging of cortical activity. *Neuron*, 26: 55-67, 2000. 90

[204] F. Babiloni, C. Babiloni, F. Carducci, L. Angelone, C. Del-Gratta, G. L. Romani, P. M. Rossini, and F. Cincotti. Linear inverse estimation of cortical sources by using high resolution EEG and fMRI priors. *Int J Bioelectromag*, 3(1), 2001. 90

[205] M. Fuchs, M. Wagner, T. Kohler, and H. Wischmann. Linear and nonlinear current density reconstructions. *J Clin Neurophysiol*, 16(3):267-295, 1999. 91

[206] P.-J. Lahaye, S. Baillet, J.-B. Poline, and L. Garnero. Fusion of simultaneous fMRI/EEG data based on the electro-metabolic coupling. In *Proc IEEE ISBI*, pages 864-867, Arlington, Virginia, Apr. 2004. 91

[207] N. J. Trujillo-Barreto, E. Martínez-Montes, L. Melie-García, and P. A. Valdés-Sosa. A symmetrical bayesian model for fMRI and EEG/MEG Neuroimage fusion. *Int J Bioelectromag*, 3(1), 2001. 94

[208] N. J. Trujillo-Barreto, E. Aubert-Vazquez, and P. A. Valdes-Sosa. Bayesian model averaging in EEG/MEG imaging. *NeuroImage*, 21(4):1300-1319, 2004. 95, 96

[209] L. Melie-García, N. J. Trujillo-Barreto, E. Martínez-Montes, T. Koenig, and P. A. Valdés-Sosa. EEG imaging via BMA with fMRI pre-defined prior model probabilities. In *Hum Brain Mapp*, Budapest, Hungary, June 2004. 96

[210] M. Sahani and S. S. Nagarajan. Reconstructing MEG sources with unknown correlations. In *Adv in Neu Info Proc Sys 16*. MIT Press, 2004. 96

[211] D. M. Schmidt, J. S. George, and C. C. Wood. Bayesian inference applied to the electromagnetic inverse problem. *Hum Brain Mapp*, 7(3):195-212, 1999. 96

[212] M. Singh, S. Kim, and T. S. Kim. Correlation between BOLD-fMRI and EEG signal changes in response to visual stimulus frequency in humans. *Magn Reson Med*, 49(1):108-114, 2003. 97

[213] R. Martuzzi, M. M. Murray, R. A. Meuli, J. P. Thiran, P. P. Maeder, C. M. Michel, R. Grave de Peralta Menendez, and S. Gonzalez Andino. Methods for determining frequency- and region- dependant relationships between estimated LFPs and BOLD responses in humans. *J Neurophysiol*, 2008. 97

[214] S. Vanhatalo, P. Tallgren, C. Becker, M. D. Holmes, J. W. Miller, K. Kaila, and J. Voipio. Scalp-recorded slow EEG responses generated in response to hemodynamic changes in the human brain. *Clin Neurophysiol*, 114(9): 1744-1754, 2003. 97

[215] S. C. Jun, B. A. Pearlmutter, and G. Nolte. MEG source localization using a MLP with a distributed output representation. *IEEE Trans Biomed Eng*, 50(6):786-9, June 2003. 101

[216] S. C. Jun and B. A. Pearlmutter. Fast robust subject-independent magnetoencephalographic source localization using an artificial neural network. *Hum Brain Mapp*, 24(1):21-34, 2005. 101

[217] S. Thaerig, N. Behne, J. Schadow, D. Lenz, H. Scheich, A. Brechmann, and C. S. Herrmann. Sound level dependence of auditory evoked potentials: simultaneous EEG recording and low-noise fMRI. *Int J Psychophysiol*, 67 (3):235-41, 2008. **xxv**, 104, 116, 118

[218] A. Roopun, M. Kramer, L. Carracedo, M. Kaiser, C. Davies, R. Traub, N. Kopell, and M. Whittington. Temporal interactions between cortical rhythms. *Frontiers in Neuroscience*, 2(2):145-154, 2008. **xxiv**, 107, 109

[219] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm. 132, 135

[220] S. Sonnenburg, G. Raetsch, C. Schaefer, and B. Schoelkopf. Large scale multiple kernel learning. *Journal of Machine Learning Research*, 7:1531-1565, 2006. 132

[221] B. Efron, H. Trevor, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32:407-499, 2004. 132

[222] R. E. Schapire. The boosting approach to machine learning: An overview. In D. D. Denison, M. H. Hansen, C. Holmes, B. Mallick, and B. Yu, editors, *Nonlinear Estimation and Classification.* Springer, New York, 2003. 132

[223] L. Pessoa and S. Padmala. Decoding near-threshold perception of fear from distributed single-trial brain activation. *Cerebral Cortex*, 17:691-701, 2007. ISSN 1047-3211. 136

[224] N. Kriegeskorte, E. Formisano, B. Sorger, and R. Goebel. Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences of the USA*, 104:20600-20605, 2007. ISSN 1091-6490. 136

[225] J.-D. Haynes and G. Rees. Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7:523-534, 2006. ISSN 1471-003X. 142

[226] J. Ramsey, S. J. Hanson, C. Hanson, Y. O. Halchenko, R. A. Poldrack, and C. Glymour. Six problems for causal inference from fMRI, Submitted. 149

[227] R. A. Poldrack, Y. O. Halchenko, and S. J. Hanson. Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychological Science*, In press. 149

[228] J.-D. Haynes. Detecting deception from neuroimaging signals–a data-driven perspective. *Trends Cogn Sci*, 12(4):126-7; author reply 127-8, Apr. 2008. ISSN 1364-6613. 151

[229] K. E. Sip, A. Roepstorff, W. McGregor, and C. D. Frith. Detecting deception: the scope and limits. *Trends Cogn Sci*, 12(2):48-53, Feb. 2008. ISSN 1364-6613. 151

[230] K. E. Sip, A. Roepstorff, W. McGregor, and C. D. Frith. Response to haynes: There's more to deception than brain activity. *Trends in Cognitive Sciences*, 12(4):127-128, 2008. ISSN 1364-6613. 151

[231] J. C. Mosher, R. M. Leahy, and P. S. Lewis. EEG and MEG: Forward solutions for inverse methods. *IEEE Trans Biomed Eng*, 46(3):245-260, Mar. 1999. 152

[232] M. A. B. Brazier. A study of the electric field at the surface of the head. *Electroencephalogr Clin Neurophysiol*, 2:38-52, 1949. 152

[233] G. W. Pruis, B. H. Gilding, and M. J. Peters. A comparison of different numerical methods for solving the forward problem in EEG and MEG. *Physiol Meas*, 14 Suppl 4A:A1-9, 1993. 153

[234] Z. Zhang. A fast method to compute surface potentials generated by dipoles within multilayer anisotropic spheres. *Phys Med Biol*, 40:335-349, May 1995. 153

[235] G. Nolte. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Phys. Med. Biol.*, 48:3637-3652, Nov. 2003. 153

[236] G. Nolte. The magnetic lead field theorem in the quasi-static approximation and its use for MEG forward calculation in realistic volume conductors. *Physics in Medicine and Biology*, 48(22):3637-52, 2004. 153

[237] 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 (2):413-97, 1993. 154

[238] S. Baillet, J. C. Mosher, and M. Leahy. Electromagnetic brain mapping. *IEEE Sig Proc Mag*, 18(6):14-30, Nov. 2001. 154, 157

[239] O. Hauk. Keep it simple: a case for using classical minimum norm estimation in the analysis of EEG and MEG data. *NeuroImage*, 21(4):1612-1621, 2004. 157

[240] S. Baillet and L. Garnero. A bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. *IEEE Trans Biomed Eng*, 44(5): 374-385, May 1997. 157

[241] R. Grave de Peralta Menendez, M. M. Murray, C. M. Michel, R. Martuzzi, and S. L. Gonzalez Andino. Electrical neuroimaging based on biophysical constraints. *NeuroImage*, 21(2):527-539, 2004. 157, 159

[242] G. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215-223, 1979. 158

[243] P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. In *SIAM Review*, volume 34, pages 561-580. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992. 158

[244] C. Phillips, M. D. Rugg, and K. J. Friston. Systematic regularization of linear inverse solutions of the EEG source localization problem. *NeuroImage*, 17 (1):287-301, 2002. 158

[245] J. Mattout, C. Phillips, W. D. Penny, M. D. Rugg, and K. J. Friston. MEG source localization under multiple constraints: an extended Bayesian framework. *NeuroImage*, 30(3):753-67, 2006. 158

[246] D. H. Brooks, G. F. Ahmad, R. S. MacLeod, and G. M. Maratos. Inverse electrocardiography by simultaneous imposition of multiple constraints. *IEEE Trans Biomed Eng*, 46(1):3-18, 1999. 158

[247] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*. Series in Automatic Computation. Prentice-Hall, Englewood Cliffs, NJ 07632, USA, 1974. ISBN 0-13-822585-0. 158

[248] B. Jeffs, R. Leahy, and M. Singh. An evaluation of methods for neuromagnetic image reconstruction. *IEEE Trans Biomed Eng*, 34(9):713-723, 1987. 158

[249] F. H. Lin, T. Witzel, S. P. Ahlfors, S. M. Stufflebeam, J. W. Belliveau, and M. S. Hamalainen. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage*, 31 (1):160-71, 2006. 158

[250] C. Phillips, M. D. Rugg, and K. J. Friston. Anatomically informed basis functions for EEG source localization: combining functional and anatomical constraints. *NeuroImage*, 16(3.1):678-695, 2002. 159

[251] G. Backus and F. Gilbert. The resolving power of gross Earth data. *Geophys J R Astron Soc*, 16:169-205, 1968. 159

[252] R. Grave de Peralta Menendez and S. L. Gonzalez Andino. A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem. *IEEE Trans Biomed Eng*, pages 440-448, 1998. 159

[253] J. Z. Wang, S. J. Williamson, and L. Kaufman. Magnetic source images determined by a lead-field analysis: the unique minimum-norm least-squares estimation. *IEEE Trans Biomed Eng*, 39(7):665-675, 1992. 159

[254] R. D. Pascual-Marqui, C. M. Michel, and D. Lehman. Low resolution electromagnetic tomography: A new method for localizing electrical activity of the brain. *Int J Psychophysiol*, 18:49-65, 1994. 159

[255] B. D. Van Veen, W. van Drongelen, M. Yuchtman, and A. Suzuki. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans Biomed Eng*, 44(9):867-880, 1997. 160

[256] S. E. Robinson and J. Vrba. Functional neuroimaging by syntheticaperture magnetometry (SAM). In T. Yoshimoto, M. Kotani, S. Kuriki, H. Karibe, and N. Nakasato, editors, *Recent Advances in Biomagnetism*, pages 302-305, Sendai, Japan, 1999. Tohoku Univ. Press. 161

[257] J. Vrba and S. E. Robinson. Differences between Synthetic Aperture Magnetometry (SAM) and linear beamformers. In J. Nenonen, R. Ilmoniemi, and T. Katila, editors, *12th Int Conf Biomagnet*, 12th International Conference on Biomagnetism, Helsinki, Finland, Aug. 2000. Biomag2000. ISBN 951-22-5402-6. 161

[258] L. J. Waldorp, H. M. Huizenga, A. Nehorai, R. P. Grasman, and P. C. Molenaar. Model selection in spatio-temporal electromagnetic source analysis. *IEEE Trans Biomed Eng*, 52(3):414-20, 2005. 162

[259] J. Malmivuo and R. Plonsey. *Bioelectromagnetism—Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, New York, 1995, 1995. **xxvii**, 166

[260] R. M. Leahy, S. Baillet, and J. C. Mosher. Integrated matlab toolbox dedicated to magnetoencephalography (MEG) and electroencephalography (EEG) data visualization and processing, 2004. 174

[261] NeuroFEM. Finite element software for fast computation of the forward solution in EEG/MEG source localisation, 2005. Max Planck Institute for Human Cognitive and Brain Sciences. 174

[262] BioPSE. Problem solving environment for modeling, simulation, and visualization of bioelectric fields, 2002. Scientific Computing and Imaging Institute (SCI). 174

[263] SCIRun. SCIRun: a scientific computing problem solving environment, 2002. Scientific Computing and Imaging Institute (SCI). 174

[264] F. Poupon. *Parcellisation systématique du cerveau en volumes d'intérêt. Le cas des structures profondes*. Phd thesis, INSA Lyon, Lyon, France, Dec. 1999. 174

[265] FreeSurfer. FreeSurfer, 2004. CorTechs and the Athinoula A. Martinos Center for Biomedical Imaging. 174

[266] D. Van Essen. Surface reconstruction by filtering and intensity transformations, 2004. 174

[267] D. Shattuck and R. Leahy. BrainSuite: an automated cortical surface identification tool. *Med Image Anal*, 6(2):129-142, 2002. 174

[268] D. Weber. EEG and MRI Matlab toolbox, 2004. 174

[269] J. E. Moran. MEG tools for Matlab software, 2005. 174

[270] A. Delorme and S. Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*, 134(1):9-21, 2004. 174

[271] F. Å. Nielsen. Bibliography of segmentation in neuroimaging, 2001. 174