Data-driven analysis for multimodal neuroimaging

vorgelegt von
Felix Bießmann

Von der Fakultät IV — Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Olaf Hellwich (Technische Universität Berlin)
1. Gutachter: Prof. Dr. Klaus-Robert Müller (Technische Universität Berlin)
2. Gutachter: Prof. Dr. Nikos K Logothetis (MPI Tübingen)
3. Gutachter: Prof. Dr. Tom Eichele (University of Bergen)

Tag der wissenschaftlichen Aussprache: 07.12.2011

Berlin, 2012

D83

In memoriam Florian Perleth

Acknowledgements

I am grateful for the support of many people during the past few years. Above all I thank my supervisors Prof. Dr. Klaus-Robert Müller, Prof. Dr. Gregor Rainer and Prof. Dr. Nikos K. Logothetis. Prof. Dr. Müller always encouraged me with scientific enthusiasm and helped in countless stimulating discussions with experienced advice on theoretical and practical aspects of statistical learning theory. With his Machine Learning Group he created an inspiringly interdisciplinary and friendly atmosphere – and gave me the opportunity to pursue research interest in neuroscience as well as in artificial intelligence in the best way I could imagine. Furthermore I am very grateful to Prof. Dr. Gregor Rainer for continuous support and for valuable insights into the experimental side of neuroscience. Of course I owe my utmost gratitude to Prof. Dr. Nikos K. Logothetis. His pioneering work on the neurophysiological basis of the hemodynamic signal is the starting point of this dissertation. I profited enormously from his scientific advice. Without his support, this dissertation would not have been possible. Also I would like to thank Dr. Yusuke Murayama for scientific advice, technical help and his interest in new analysis methods. Moreover I am indebted to the outstanding experimental expertise of Axel Öltermann, Dr. Marc Augath, Dr. Jozien Goense, Dr. Alexander Rauch, Prof. Dr. Robert Kretz, Jule Veit and Anwesha Bhattacharyya. By the same token I would like to thank Dr. Arthur Gretton, Dr. Jakob Macke and Prof. Dr. Matthias Bethge for their help on mathematical problems during my time in Tübingen. I also thank Dr. Andreas Harth for an interesting research project outside of neuroscience. I was lucky to meet many friendly, creative and bright minds who helped a lot for calibrating my sense of whats right and wrong in science and beyond, among them Michael Gaebler, Dr. David Greenberg, Daniel Lebrecht, Dr. Dr. Franz Kiraly , Stefan Haufe, Paul von Bünau and Jule Veit. Special thanks go to Frank C. Meinecke for constant and patient supervision. His experienced skepticism, his sharp thinking and his talent to explain complex concepts in a comprehensive way were invaluable to me. To him I also owe the layout of this thesis. Also I would like to thank Valentina Mosienko for her patience and support. Finally I owe my deepest gratitude to my family. The most important decisions which led to this thesis were not taken by me, but by my parents. Thank you for making me curious.

Abstract

Neuroimaging, the measurement, analysis and visualization of neural activity, contributed considerably to our understanding of information processing in the brain. The availability of non-invasive neuroimaging devices such as *functional magnetic resonance imaging* (fMRI) has been increasing rapidly throughout the last two decades. Nowadays every interested student can obtain non-invasively high resolution image time series of the *blood-oxygen level dependent* (BOLD) signal using fMRI. How exactly neural activity is reflected in the BOLD contrast is still subject of active research. For a more accurate interpretation of the fMRI signal and the underlying neurovascular coupling mechanisms, combined measurements of intracranial neural activity and fMRI signals are indispensable. Such simultaneous measurements have become technically possible – however appropriate analysis methods are still lacking. Classical analysis approaches rely on simplifying assumptions about the neurovascular coupling dynamics. These assumptions are convenient but numerous studies have provided empirical evidence against them.

In this dissertation a novel analysis method, termed *temporal kernel Canonical Correlation Analysis* (tkCCA), will be developed, tested on artificial data and applied to experimental data in order to investigate neurovascular coupling mechanisms. TkCCA estimates dependency structures between high dimensional data with complex temporal coupling dynamics. The important advantages of tkCCA compared to standard methods are a) tkCCA can be directly applied to multimodal data, b) tkCCA is very efficient for high dimensional data with few data points (as is the case for fMRI) and c) tkCCA does not make use of restrictive assumptions about the data generating process. In particular tkCCA can be used to analyze high dimensional simultaneous measurements of neural activity and fMRI signals. Predictions of neural activity using tkCCA are better than when using classical methods. Basic research as well as clinical applications can profit from this more accurate prediction. Besides tkCCA is readily applicable to other domains in which data streams have high dimensional features that are non-instantaneously coupled, such as data from social networks in the World Wide Web.

Zusammenfassung

Bildgebende Verfahren in den Neurowissenschaften haben unser Verständnis von Informationsverarbeitung im Hirn entscheidend geprägt. Die schnelle Verbreitung von nicht-invasiven bildgebenden Verfahren wie etwa der *funktionellen Magnet-Resonanz Tomographie* (fMRT) in den letzten beiden Dekaden erlaubt es heutzutage jedem interessierten Studenten nicht-invasiv den Blutsauerstoffgehalt im Hirn zu messen und so räumlich hochaufgelöste Zeitreihen von Hirnaktivität aufzuzeichnen. Wie genau sich jedoch neuronale Aktivität im *Blut-Sauerstoff Gehalt abhängigem Kontrast* (englisch: *Blood-oxygen level dependent* oder BOLD contrast) wiederspiegelt, ist trotz intensiver Forschung an der *Neurovaskulären Kopplung* nach wie vor aktuelles Forschungsthema. Für ein besseres Verständnis des fMRT Signals sind kombinierte Messungen von intrakranialer neuronaler Aktivität und fMRT Signalen zwingend erforderlich. Diese *multimodalen* Simultanmessungen sind inzwischen technisch möglich. Jedoch fehlen geeignete Analysemethoden. Gängige Verfahren basieren auf vereinfachenden Annahmen über die neurovaskuäre Kopplungsdynamik. Diese Annahmen sind zwar praktisch, erwiesen sich aber in zahlreichen Studien als falsch.

In dieser Dissertation wird ein neuartiges Analyseverfahren, *temporal kernel Canonical Correlation Analysis* (tkCCA), entwickelt, getested und angewandt. TkCCA schätzt Abhängigkeitsstrukturen in hochdimensionalen Daten mit nicht-instantaner Kopplung. Die entscheidenden Vorteile von tkCCA gegenüber herkömmlichen Methoden sind a) direkte Anwendbarkeit auf multimodale Daten, b) Effizienz bei hochdimensionalen Daten mit wenig Datenpunkten und c) Verzicht auf einschränkende Annahmen über die generativen Modelle der gemessenen Daten. Insbesondere erlaubt tkCCA die Analyse hochdimensionaler *multimodaler* Simultanmessungen von neuronaler Aktivität und fMRT Signalen. Mit Hilfe von tkCCA können neuronale Signale besser aus multivariaten fMRI Messungen vorhergesagt werden. Davon können sowohl Grundlagenforschung als auch klinische Diagnostik profitieren. Darüberhinaus ist tkCCA direkt anwendbar auf andere Arten hochdimensionaler Datenströme mit nicht-instantanen Abhängigkeiten, wie etwa in sozialen Netzwerken im World Wide Web.

x

Contents

1 Introduction 1

I Unimodal Neuroimaging

2 A short history of neuroimaging 9

3 Unimodal neuroimaging 11

4 Unimodal analysis approaches 19

II Multimodal Neuroimaging

5 Multimodal neuroimaging 35

6 Multimodal analysis approaches 45

7 Data-driven multimodal analysis 53

8 Model-free analysis of neurovascular coupling 73

9 Summary and outlook 103

Appendix

A Mathematical preliminaries 109

B Experimental protocols 113

Chapter 1

Introduction

> … it is in the brain that everything takes place.
>
> Oscar Wilde

The brain is an exceptionally interesting organ. Like no other organ it is in the focus of attention of a broad spectrum of scientific disciplines, ranging from natural science over psychology to philosophy. Many of the insights we have about the brain were gained by measuring its electrical and metabolic activity. Measurement of brain activity and the visualization thereof using appropriate analysis techniques is called *neuroimaging*. Various neuroimaging techniques have been developed, but each method has its specific technological and physiological limits. Some techniques, such as *electroencephalography* (EEG), can visualize brain activity at a high temporal resolution, but only at a very poor spatial resolution. Other methods, such as measurements of the *blood-oxygen level dependent* (BOLD) signal for instance, can image the whole brain at once, in particular deep structures that cannot be measured with EEG, but they cannot give insights in the fast temporal dynamics of brain activity. This is why combinations of multiple neuroimaging modalities have become popular. Multimodal imaging setups can take advantage of complementary views on neural activity and enhance our understanding about how neural information processing is reflected in each modality. In order to exploit the potential of multimodal methods, dedicated analysis methods are needed. Many solutions to this data integration problem have been proposed. However the complex physiological processes that give rise to the BOLD contrast are difficult to model. Thus, every analysis method for hemodynamic neuroimaging data has to make simplifying assumptions about the data generating process. This allows for more efficient computations and easier interpretation of the results. However due to these assumptions some aspects of brain activity might be lost in the analysis. In this dissertation a novel analysis framework for multimodal neuroimaging data will be proposed. Applying this analysis method on multimodal neural data, a simplifying assumption underlying many studies based on BOLD contrast will be tested empirically.

Measurement, analysis and visualization of brain activity is called neuroimaging

The relationship between brain activity and its hemodynamic response is called neurovascular coupling

What is measured with fMRI? One of the most important applications of multimodal neuroimaging is the investigation of the *neurovascular coupling mechanisms*, the relationship between brain activity and the hemodynamic signal. While accurate measurements of brain activity require intracranial electrodes, hemodynamic signals can be measured non-invasively using for instance fMRI [Ogawa et al., 1990]. The hope that non-invasive techniques will eventually replace invasive neural recordings is the basis for the enormous success of fMRI in recent years [Friston, 2009]. However the exact relationship between neural activity and fMRI signals is still subject of active research [Logothetis, 2008]. The physiological processes underlying the generation of the BOLD signal are still not sufficiently well understood to model them accurately. Many detailed models have been proposed, but they are difficult to test empirically. A few reasons for this are: Recording the quantities involved in neurovascular coupling is technically challenging. Large scale recordings of all electrical and chemical forces involved cannot be obtained yet. Another reason is the computational complexity of the analysis. Classical statistical methods are not designed for state of the art multimodal data with tens of thousands of dimensions and only a few data points – a classical problem faced by researchers working on fMRI data. Technical advances in the field of multimodal recordings enable researchers to obtain high resolution hemodynamic measurements simultaneously with intracranial neurophysiological recordings [Logothetis et al., 2001, Oeltermann et al., 2007, Goense and Logothetis, 2008]. For these data, novel analysis methods are needed [Dale and Halgren, 2001, Friston, 2009].

A novel analysis framework In this dissertation a novel analysis framework termed temporal kernel canonical correlation analysis (tkCCA) is proposed. It is tailored to the specific needs in multimodal neuroimaging: Scalability to high dimensional data and the ability to account for arbitrary non-instantaneous coupling mechanisms while making only minimal assumptions about the data generating process and the coupling mechanisms. In order to meet these requirements, tkCCA combines well established statistical learning techniques with modern machine learning methods.

Does spatiotemporal variability of fMRI signals contain *neural* information? After introduction of tkCCA and validation thereof on synthetic data, the method is used to test a simplifying assumption underlying many neuroimaging methods, in particular multimodal analyses. Most neuroimaging methods assume that the spatial dynamics of the hemodynamic response to neural activation is separable from the temporal dynamics of the response. Non-separable spatiotemporal variability of the hemodynamic response is thus neglected in these methods. Empirical evidence suggests that this spatiotemporal separability assumption is a good approximation:

When considering the voxels in an fMRI image sequence located around a neural ensemble of interest after that ensemble was stimulated, the temporal dynamics are very similar. However there is substantial evidence showing that the hemodynamic response varies across brain regions [Aguirre et al., 1998] or different cortical layers [Yacoub et al., 2006]. Taking this variance into account can reveal areas of activation that would have been overlooked by methods that assume spatiotemporal separability [Mourão-Miranda et al., 2007, Lu et al., 2007]. The scientific hypothesis underlying the spatiotemporal separability assumption is: Does the spatiotemporal variability of the hemodynamic response carry information about neural signals? While the above studies provide only indirect evidence, tkCCA applied to simultaneous recordings of neural and hemodynamic activity can directly test this hypothesis.

Overview of this dissertation

The structure of this thesis is divided in two parts. The first part will give a short overview over popular neuroimaging methods in chapters 2, 3 and 4. The second part will deal with the combination of neuroimaging methods in multimodal setups. Chapter 5 gives an introduction to multimodal neuroimaging setups and some motivating examples of their application in basic neuroscientific research and clinical application. Common analysis standards established in the literature will be reviewed in chapter 6.

The main part of this thesis is chapter 7 in which the proposed algorithm is developed. In collaboration with the Max-Planck Institute for Biological Cybernetics, Tübingen, we applied tkCCA for estimation of neurovascular coupling dynamics and high accuracy prediction of neural activity from simultaneously recorded fMRI signals. Chapter 8 will show applications of the proposed algorithm on artificial and real data.

Own contributions

The algorithm and its application to multimodal neural data have been published in internationally renown journals in the machine learning and neuroimaging community; a detailed overview is given on the next page. Early stages of this work have been presented in form of conference abstracts at the Forum of the European Neuroscience Society (FENS) in Geneva [Bießmann et al., 2008], the Computational Neuroscience Society (CNS) in Berlin [Bießmann et al., 2009] and the Computational and Systems Neuroscience (COSYNE) Conference in Salt Lake City [Bießmann et al., 2010a]. This work profited from scientific exchange after presentation of preliminary results at the RIKEN Brain Science Institute (Tokyo), the Cognitive

and Neurobiological Imaging Lab at Stanford University and the Redwood Center for Theoretical Neuroscience at University of California, Berkeley. Parts of this thesis are based on a review article on multimodal neuroimaging analysis methods [Bießmann et al., 2011b]. Next to the work on multimodal neuroimaging we also explored applications in other domains such data mining [Bießmann and Harth, 2010]. A summary of the work that is published as peer reviewed manuscripts is given in the following.

Peer reviewed manuscripts

1. Bießmann, Meinecke, Gretton, Rauch, Rainer, Logothetis, and Müller, *Temporal Kernel CCA and its Application in Multimodal Neuronal Data Analysis*, **Machine Learning**, 2010

   Summary Temporal kernel canonical correlation analysis (tkCCA) was proposed and tested in simulations and preliminary data from simultaneous measurements of neurophysiological recordings and fMRI signals during sensory stimulation.

   Contribution I developed parts of the recording software, did all analyses and wrote the manuscript.

2. Murayama, Bießmann, Meinecke, Müller, Augath, Oeltermann, and Logothetis, *Relationship between neural and hemodynamic signals during spontaneous activity studied with temporal kernel CCA*, **Magnetic Resonance Imaging**, 2010

   Summary In this manuscript we showed that tkCCA can robustly estimate neurovascular coupling dynamics from recordings of spontaneous activity in primary visual cortex.

   Contribution I did parts of the analyses and wrote parts of the manuscript.

3. Bießmann and Harth, *Analysing Dependency Dynamics in Web Data*, **Proceedings of AAAI Symposium**, 2010

   Summary An alternative application of tkCCA is illustrated on social network data; using time resolved listening behavior of users in my friend subgraph on http://last.fm I extracted music trends and identified users who where ahead of this trend and those who lagged behind.

   Contribution I collected the data, analyzed it and wrote the manuscript.

4. Bießmann, Plis, Meinecke, Eichele, and Müller, *Analysis of Multimodal Neuroimaging Data*, **IEEE Reviews in Biomedical Engineering**, 2011

Summary This review summarizes the state of the art in multimodal neuroimaging analysis with a special focus on data driven methods.

Contribution I wrote major parts of the manuscript and provided application examples from artificial and real data.

5. Bießmann, Murayama, Logothetis, Müller, and Meinecke, *Improved Decoding of Neural Activity from fMRI Signals: Towards Non-Separable Spatiotemporal Deconvolutions*, in revision

   Summary In this manuscript we show that the spatiotemporal variability of the hemodynamic response carries information about neural activity that most fMRI analysis methods neglect.

   Contribution I analyzed the data and wrote the paper.

6. Bhattacharyya, Bießmann, Veit, Kretz, and Rainer, *Functional and laminar dissociations between muscarinic and nicotinic cholinergic neuromodulation in the tree shrew primary visual cortex*, to appear in European Journal of Neuroscience

   Summary In this paper we showed differential effects of the neuromodulator Acetylcholine (ACh) on visual information processing in different layers of primary visual cortex.

   Contribution I developed parts of the stimulus presentation software, helped recording some of the data, contributed to the data analysis and wrote minor parts of the manuscript.

7. Rauch, Zhang, Bießmann, Meinecke, Goense, Müller, Rainer, and Logothetis, *Baseline BOLD signal shift in macaque primary visual cortex (V1) after local application of Acetylcholine*, in preparation

   Summary In this paper we showed differential effects of the neuromodulator Acetylcholine (ACh) on visual information processing as measured with neurophysiological recordings and hemodynamic activity. The BOLD signal exhibited a pronounced increase in baseline activity, while the effects on neural activity were rather heterogeneous. A series of additional experiments outside of the scanner helped to resolve the heterogeneous effects on neural activity: In [Bhattacharyya et al., 2011] we could show that the effects of ACh can be dissociated with respect to laminar position and receptor type.

   Contribution I developed the realtime recording system, helped recording some of the data, performed all analyses and wrote minor parts of the manuscript.

Part I

Unimodal Neuroimaging

Chapter 2

A short history of neuroimaging

> Es war von vornherein zu erwarten, daß auch im Zentralnervensystem […] bioelektrische Erscheinungen nachweisbar seien.
>
> Berger [1929], *Über das Elektroenkephalogramm des Menschen*

Since the late 19th century researchers explore how cognitive functions are reflected in measurements of neural activity [Caton, 1875]. Using different neuroimaging techniques, disciplines such as psychology, biology and medicine accumulated knowledge about how what we perceive, feel, think and do is related to the complex activity patterns of neurons in our central nervous system. Table 2.1 highlights a selection of important achievements in the history of (multimodal) neuroimaging. Animal studies have been indispensable for a better understanding of how neural activity is related to the outside world, some examples are [Mountcastle, 1957, Hubel and Wiesel, 1959]. But the most fascinating mental phenomena, such as higher cognitive functions like language or reasoning, are difficult to study with animals. Often long training phases are needed to establish an experimental paradigm. Humans have a clear advantage here: We can use language to communicate what a subject is supposed to do in an experiment. Human neuroscientific experiments rely on non-invasive measurements of brain activity. They can be obtained for instance by electroencephalography (EEG) [Berger, 1929], magnetoencephalography (MEG) [Cohen, 1968], near infrared spectroscopy (NIRS) [Jöbsis, 1977] or functional magnetic resonance imaging (fMRI) [Ogawa et al., 1990]. The quantities measured by each of these modalities have different physiological origins and thus different limitations and advantages. And each modality reflects neural activity at a different spatiotemporal scale. A short summary of various neuroimaging techniques is given in the next chapter.

1875	Caton	Electrocorticography (ECoG) reveals electrical activity in the brain in response to visual stimulation
1929	Berger	Electroencephalography (EEG) reveals brain oscillations between $8-12$Hz (α-rhythms) in occipital areas associated with states of vigilance
1957	Mountcastle	Detailed topographic mapping of sensory modalities using intracranial electrophysiology
1968	Cohen	Magnetoencephalography (MEG) for measuring α activity in occipital cortex
1977	Jöbsis	Near infrared spectroscopy (NIRS) for non-invasive imaging of brain activity
1986	Grinvald et al.	Combined intrinsic optical signal imaging and intracranial electrophysiology to investigate neurovascular coupling
1990	Ogawa et al.	Functional magnetic resonance imaging (fMRI) to visualize brain activity
2000	Dale et al.	Sequential fMRI and MEG
2001	Lemieux et al.	Simultaneous fMRI and EEG
2001	Logothetis et al.	Simultaneous fMRI and intracranial microelectrode recordings

Table 2.1: A short history of (multimodal) neuroimaging methods

Chapter 3

Unimodal neuroimaging

Most neuroimaging modalities measure either electrophysiological or hemodynamic signals. A widely used neuroimaging technique for electrophysiological activity with an exquisite spatial and temporal resolution are intracranial microelectrode recordings. Among hemodynamic modalities, fMRI became most popular [Friston, 2009]. Electrophysiological recordings pick up changes in electromagnetic fields induced by neural activity; energy consumption of neural activity is correlated with *blood oxygenation*. The hemodynamic signal thus reflects neural activity indirectly. Both electrophysiological and hemodynamic signals can be measured invasively and non-invasively. Electrodes can be placed directly in the neural tissue, on the cortical surface (electrocorticograms or ECoG) [Caton, 1875] or on the skull (electroencephalograms or EEG) [Berger, 1929]. Neural activity is also reflected in the magnetic field fluctuations which can be measured non-invasively by magnetoencephalograms (MEG) [Cohen, 1968]. EEG and MEG measure changes in electrical and magnetic fields, respectively, on the scalp surface. The origin of EEG and MEG signals are electrical dipoles in the cortex emerging from synchronized activity of neighboring neurons with elongated shape [Murakami and Okada, 2006]. The strongest signal in EEG recordings arises from dipoles oriented perpendicular to the scalp surface [Nunez and Srinivasan, 2006]. MEG in contrast is most sensitive to cortical dipoles tangential to the scalp [da Silva and Niedermayer, 1998]. The magnetic fields measured with MEG are not affected by volume conduction and can resolve finer structures than EEG [Grynszpan and Geselowitz, 1973, Cuffin and Cohen, 1979]. Hemodynamic activity can be measured semi-invasively using intrinsic optical imaging [Grinvald et al., 1986] or non-invasively by functional magnetic resonance imaging (fMRI) [Ogawa et al., 1990] and near infrared spectroscopy (NIRS) [Jöbsis, 1977]. In NIRS infrared light is sent through the scull and the cortical tissue; as oxygenated blood has different wavelength absorption characteristics than deoxygenated blood, neural activity can be measured by examining the reflected light. The spatial resolution is much lower than fMRI and imaging is restricted to the cortical surface but the temporal resolution can be higher than fMRI; besides the setup is much cheaper and less complex than that of fMRI. Next to these non-invasive neuroimaging techniques, there are semi-invasive techniques that require opening the

Neural activity is reflected directly in electrical field changes and indirectly in the amount of oxygen bound to hemoglobin molecules; both can be measured invasively and non-invasively;

Measurement	Origin	Resolution		Invasive
		Spatial	Temporal	
Electrophysiological activity				
Intracranial recordings	single cell / population activity	high	high	yes
EEG	cortical dipoles *orthogonal* to scalp	low	high	no
ECoG	cortical dipoles *orthogonal* to scalp	medium	high	yes
MEG	cortical dipoles *tangential* to scalp	low	high	no
Hemodynamic activity				
Optical imaging	blood oxygenation and volume	high	high	yes
NIRS	blood oxygenation and volume	low	medium	no
fMRI	blood oxygenation	high	low	no

Table 3.1: Simplified overview of popular neuroimaging modalities;

skull but no penetration of neural tissue as in the case of intracranial electrophysiological recordings. Semi-invasive preparations such as ECoG and intrinsic optical imaging offer a higher resolution than completely non-invasive methods and bear fewer risks than invasive recordings. ECoG measures electrical oscillations between electrodes directly on the cortex. Similarly hemodynamic activity can be imaged with optical imaging in minimally invasive preparations also in humans [Arthur and Nader, 2002].

3.1 Electrophysiogical measurements

The computations carried out by our brain are reflected in changes of electrical potentials across the cell membrane of neurons [Hodgkin and Huxley, 1952]. Understanding neural computations requires measurements of electrophysiological activity. This can be done at various levels, invasively, semi-invasively with ECoG or non-invasively with EEG or MEG. The most detailed measurements of neural activity can be obtained with invasive electrophysiological recordings.

Physiological origin Invasive electrophysiological recordings measure the changes in electrical fields emerging from neural activity. Neurons communicate mainly via activation of chemosensitive ion channels located on the (post-)synapse, illustrated as black dots in figure 3.1. If a neuron releases neurotransmitters at the synapse, ion channels on the postsynaptic neuron open. The opening allows ions to flow down their electrochemical gradient. This results in a depolarization of the dendritic part of the cell with respect to the extracellular medium (indicated by red plus sign in

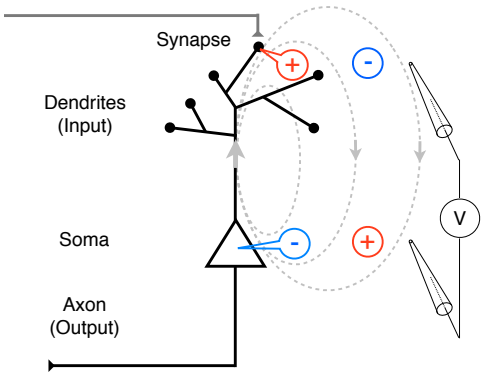

Figure 3.1: Sketch of a neuron: In grey the axon of another cell forming a synapse with the neuron in black. During rest, the intracellular medium is approximately at -70mV with respect to the extracellular medium; neurotransmitters released at the synapse result in a depolarization in the dendrite, the cell becomes an electrical dipole: the soma is negatively charged relative to the dendrite; if many neurons are arranged in parallel and receive synchronized dendritic input, this dipole will become stronger – eventually strong enough give rise to electromagnetic fields that can be measured outside of the scull using EEG;

the post synapse and a blue minus sign in the extracellular medium). Synchronized depolarization of large ensembles of neighboring neurons result in the generation of electromagnetic dipoles. The fields emerging from these dipoles can be measured extracellularly with electrodes, on the cortical surface with ECoG or outside the skull with non-invasive EEG or MEG measurements. The strength of the measured signal depends critically on the dipole generating neurons, their shape and their spatial arrangement in an ensemble. Aspiny inhibitory interneurons with radially symmetric dendritic trees around their somata for instance will form dipoles that are difficult to measure in extracellular recordings. Other cells, such as large pyramidal neurons, have a bipolar shape with the some on one end and the dendrite elongating in parallel to neighboring neurons (see fig. 3.1). These pyramidal ensembles generate dipoles with are much stronger in electrophysiological recordings [Murakami and Okada, 2006]. The strongest signal in EEG recordings arises from dipoles oriented perpendicular to the scalp surface, the strongest signal in MEG recordings are dipoles tangentially to the scalp [da Silva and Niedermayer, 1998, Nunez and Srinivasan, 2006]. Invasive electrical recordings offer the highest spatiotemporal resolution. Two aspects are differentiated in intracranial electrophysiological recordings, fast discharges – spikes or action potentials (APs) – and low frequency content, often called *local field potentials* or LFP. Neuronal spikes are associated with the output of the computations carried out by a neuron. The high frequency spectrum of neurophysiological recordings containing the spiking activity of many single units is called *multi-unit activity* (MUA). As the electrode is at a fixed position relative to

Local Field Potentials *(LFPs) are slow oscillations in neurophysiological signals; fast (above 1KHz) oscillations are called* multi-unit activity *(MUA)*

the cells in its surround, action potentials of different cells can be differentiated by their distinct AP shapes recorded at the electrode. Classifying single cell units is called *spike sorting*. The LFP is hypothesized to reflect subthreshold membrane oscillations (the input to or state of a neuron before it is sending out spikes). Metabolic signals of brain activity, such as the BOLD contrast, has been reported to be more correlated with LFPs than spikes [Logothetis et al., 2001, Goense and Logothetis, 2008]. This finding is consistent with the larger number of mitochondria[1] found in the dendritic parts of neurons (the input site) as compared to the axons (the output part) [Wong-Riley, 1989, Attwell and Laughlin, 2001].

Signal properties and limitations Electrophysiological measurements have a high spatiotemporal resolution. Intracranial recordings have the potential to resolve single cell activity. However the effective spatial resolution depends on the number and arrangement of electrodes. For intracranial neurophysiology the temporal resolution is typically around 20Khz. Depending on the number of electrodes up to 500 single cells can be recorded at once [Lehev and Nicolelis, 2008]. Recently there has been increased interest in LFP signals. They are easy to measure, with much lower sampling rates than spikes, but exhibit a similar sensitivity for sensory features compared to spikes in early visual cortex [Xing et al., 2009]. Also the LFP is shown to correlate better than spikes with the fMRI signal [Logothetis et al., 2001, Goense and Logothetis, 2008]; for many applications the high temporal resolution of spikes is not needed and LFPs represent a promising alternative [Waldert et al., 2009]. The spatial resolution of LFPs is on the order of hundreds of micrometers [Katzner et al., 2009]. Non-invasive electrophysiological measures like EEG have a lower spatial resolution on the order of several millimeters as measured by comparing EEG based dipole estimates with fMRI activation centers [Im et al., 2007]. Although there is evidence that the phase of the EEG oscillations is coupled to intracranial spikes [Whittingstall and Logothetis, 2009] neuronal spikes cannot be detected directly using EEG. Hence EEG as well as MEG is typically sampled at frequencies below 1KHz.

3.2 Hemodynamic measurements

Neural activity is reflected indirectly in the hemodynamic response. Hemodynamic activity in the entire brain can be measured non-invasively using fMRI. Nowadays commercially available fMRI setups can easily be operated without any knowledge about MRI physics. Thus fMRI has become by far the most popular neuroimaging technique [Friston, 2009]. It is important to keep in mind that fMRI does not measure neural activity directly. The relationship between neural and hemodynamic

[1] *Mitochondria* are the cellular power plants and provide the energy needed for the ion pumps preserving electrochemical gradients across neuronal membranes.

signals is still subject of active research [Logothetis, 2008]. As more and more non-technically minded researchers are making use of this powerful technique, recent studies highlighted the importance of solid statistical standards for fMRI analysis [Kriegeskorte et al., 2009, Bennett et al., 2009, Vul et al., 2008].

Physiological origin Neural activity in the brain consumes energy which is delivered by the blood stream. The blood supply is controlled by the highly complex cortical microvasculature, shown in fig. 3.2 as a *vascular corrosion cast*[2] of primary visual cortex of the macaque monkey. Most importantly the vascular system has to ensure that there is always enough oxygen delivered to the neurons. A blockage of the arterial vessels on the cortical surface can affect whole cortical columns underneath and thus can have devastating consequences on cognitive functions. Oxygen is carried through the blood stream via erythrocytes, the red blood cells. The cell plasma of erythrocytes is rich in hemoglobin molecules, which transport oxygen. The important aspect for neuroimaging is that deoxygenated hemoglobin (HbR) and oxygenated hemoglobin (HbO) have different magnetic and light absorption properties. This gives rise to the blood-oxygen level dependent (BOLD) signal. The BOLD contrast is a complex combination of blood oxygenation, blood flow and blood volume [Buxton et al., 2004]. Hemodynamic signals can be measured invasively using *intrinsic signal optical imaging* (ISOI) [Grinvald et al., 1986, Frostig et al., 1990] as well as non-invasively using functional magnetic resonance imaging (fMRI) [Ogawa et al., 1990] or *near infrared spectroscopy* (NIRS) [Jöbsis, 1977].

Oxygenated and deoxygenated blood have different magnetic and light absorption properties

Signal Properties and Limitations The temporal resolution of intrinsic optical imaging data is typically around 10Hz (see e.g. [Berwick et al., 2008]). While optical imaging can visualize hemodynamic signals on the cortical surface at a high resolution, the depth resolution is rather poor. Intrinsic optical imaging requires opening (or at least removing parts of) the skull. Although initial applications of optical imaging were in basic neuroscience research [Grinvald et al., 1986, Frostig et al., 1990] its usefulness for diagnostic purposes is also being explored in human patients [Pouratian et al., 2002]. NIRS operates with infrared light which can travel through the intact skull, measurements can thus be taken non-invasively. This non-invasiveness comes at the price of a poor spatial resolution. The advantage is that NIRS setups are simple, low-cost and portable [Villringer and Chance, 1997, Wolf et al., 2007]. The advantages of optical imaging and NIRS are combined in fMRI: Like NIRS, fMRI measurements are non-invasive. And like optical imaging, the spatial resolution of fMRI measurements is high; in contrast to optical imaging, fMRI can image deep subcortical structures. Using specialized imaging protocols and hardware, fMRI can resolve cortical laminae [Goense and Logothetis, 2006]. The

[2] In *vascular corrosion casts* vessels are perfused with plastic and surrounding tissue is removed; the negative of the vascular system can be imaged at high resolution with electron microscopy;

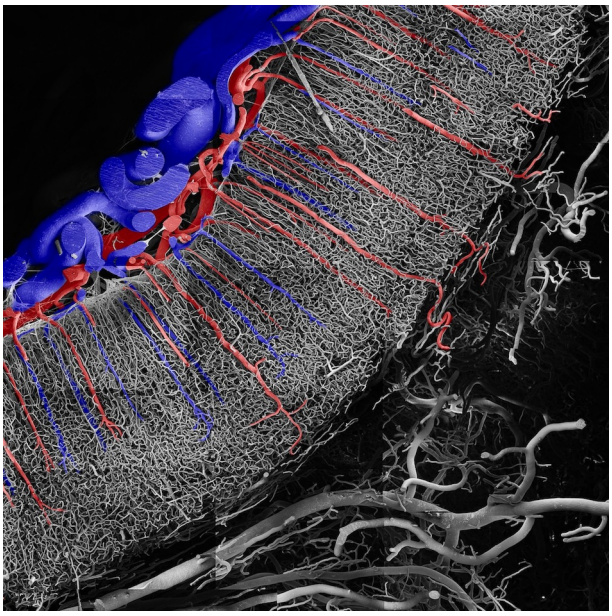

Figure 3.2: Vascular system in primary visual cortex of the macaque monkey; veins are shown in blue, arteries in red, micro vessels in grey (taken from [Keller et al., 2008] with kind permission of Anna Lena Keller, MPI Biological Cybernetics, Tübingen)

Cortical microvasculature gives rise to complex hemodynamic activation patterns

point spread function of fMRI signals is on the order of 2mm [Shmuel et al., 2007, Sirotin et al., 2009]. Of course the neural signal is blurred by the vascular structure, but this can actually be helpful for decoding the neural signal as the microvasculature has a functional meaning. Recent work demonstrates a laminar specificity of the ratio of microvasculature-to-cell-density [Weber et al., 2008, Tian et al., 2010]. Incorporating these insights into detailed models of neurovascular coupling allows for a more detailed analysis of fMRI signals [Boas et al., 2008, Guibert et al., 2010]. The temporal resolution of fMRI can be below 1Hz, but high temporal resolution comes at the price of poor signal to noise ratio of the image sequence. In human experiments the spatial resolution of fMRI signals is typically much lower and the temporal resolution is usually below 1Hz; an advantage of human fMRI is that single subjects' brain scans can be co-registered with template brains. This makes multi-subject analysis possible. But the co-registration requires spatial smoothing which reduces the effective spatial resolution of human fMRI scans. To summarize, the best spatiotemporal resolution for whole brain imaging is obtained by fMRI. As fMRI requires hardware which is expensive to maintain and is not portable, optical imaging methods are a sensible alternative for measuring hemodynamic activity in many applications.

3.3 Primary visual cortex: A well studied brain region

Neurons in primary visual cortex (V1) respond selectively to bars of light at a certain orientation

When investigating neurovascular coupling, it is reasonable to choose a brain region that is well characterized. Testing hypotheses about how multiple modalities are related is easier if the functional and anatomical properties of the brain region measured are well understood. In terms of the functional characterization the best studied sensory modality is undoubtedly the visual sense. In their seminal work [Hubel and Wiesel, 1959] found that cortical neurons in occipital regions respond very selectively to bars of light of a certain orientation when presented in the visual field. In 1981 the authors received the Nobel Prize. As the primary visual cortex (also called V1) is one of the best understood brain areas we will focus on data recorded in this brain region throughout the entire dissertation. An example of the cortical functions first described in [Hubel and Wiesel, 1959] is shown in figure 3.3. The data was taken from [Bhattacharyya et al., 2011]. When placing a microelectrode in V1 and presenting drifting gratings of different orientation (see fig. 3.3A) to the animal, cortical neurons will emit more action potentials than during rest. Cells in V1 are selective for gratings of a certain orientation: A cell will emit more spikes when the *preferred orientation* was shown as opposed to when the *orthogonal orientation* is shown (see fig. 3.3B).

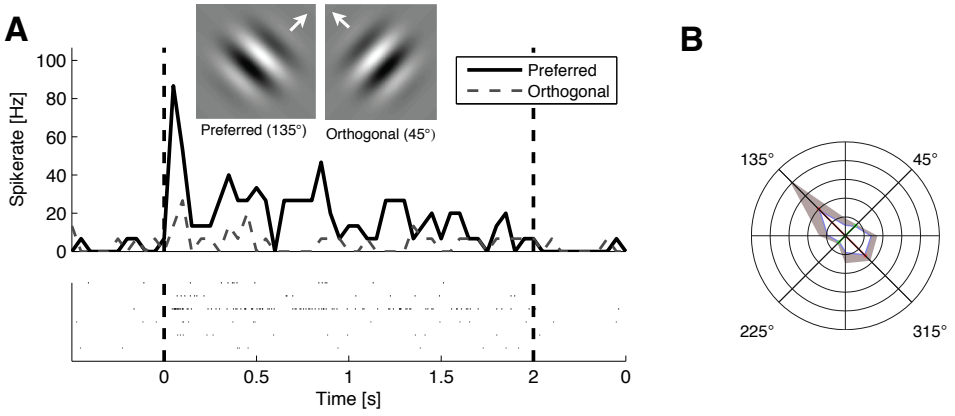

Figure 3.3: Selectivity of neurons recorded in primary visual cortex in response to moving gratings of different orientation; **A**: Peristimulus time histogram (top panel) and single trial spike raster plots (bottom panel) of a cortical cell's spiking response to a preferred stimulus (grating with 135° angle) and orthogonal stimulus (45°); stimulus on- and offset is marked by vertical black dashed line; highest spike rates were recorded when the *preferred orientation* was shown, lowest spike rates when the *orthogonal orientation* was shown; **B**: Polar plot summarizing the responses (in a 2s window) to all stimulus orientations (in degree); black circles indicate 10Hz increments in spike rate, a light blue inner polygon denotes median firing rate for a given orientation, a grey tube indicates variance across trials;

Chapter 4

Unimodal analysis approaches

This chapter will recapitulate popular analysis concepts for unimodal neuroimaging data in order to set the stage for multimodal extensions in later chapters. From a data analyst's point of view, unimodal analysis methods can be categorized into three classes:

- Supervised methods
  Are concerned with *regression* between measurements and/or experimentally controlled variables.

- Unsupervised methods
  Find structure in measurements in an *explorative* fashion.

- Model driven methods
  Fit *physiological models* to measurements.

Many neuroimaging studies combine methods from more than one of these three classes. Thus the analysis of a particular neuroimaging study is often difficult to assign to one of these three classes. Nonetheless this categorization can be helpful. Each of these classes has advantages and drawbacks. Not all methods can be applied to every data set. For instance supervised methods usually require an experimentally controlled variable as regressor, which is not available in every experimental setting. And some scientific questions can only be addressed with methods from one of the three. For example in order to test a scientific hypothesis about a certain biological process it is often convenient to formulate a model and compare the model predictions with experimental data. While an exhaustive review of these categories is beyond the scope of this dissertation, a basic understanding of standard neuroimaging analyses will be helpful in later chapters. The following sections will give a short introduction to some methods popular in classical unimodal neuroimaging analyses.

4.1 Supervised Methods

Supervised methods have in common that a regressor $x \in \mathbb{R}^U$ is used to explain a target variable $y \in \mathbb{R}^1$

$$f(x) = \hat{y} + \varepsilon \tag{4.1}$$

where $f(.)$ denotes a function that maps the regressor onto an estimate $\hat{y}$ of the target variable y and ε denotes noise variance that is not captured by $f(.)$. This function is chosen such that some difference measure between target variable and its estimate is minimized

$$\underset{f(.)}{\operatorname{argmin}} \left(\|\hat{y} - y\|_p \right), \tag{4.2}$$

where the distance measure often is the p norm $\|x\|_p = \sqrt[p]{\sum x^p}$. Typically one choses $p = 2$ such that eq. (4.2) minimizes the *least-squares error* [Legendre, 1805]. During the last decades two main streams of supervised neuroimaging data analysis have emerged, mass-univariate methods, also known as statistical parametric maps (SPMs) [Friston and Buchel, 2003], and multivariate pattern analysis (MPA) [Cox and Savoy, 2003, Haynes and Rees, 2006, Norman et al., 2006]. Both are based on (mostly linear) regression, only the role of target variables and regressors are different. There is an ongoing debate as to whether mass-univariate SPMs are better than MPA approaches; the main argument for SPMs is that brain activation can be better localized with mass-univariate methods [Kiebel and Friston, 2004]; other authors advocate the higher sensitivity of multivariate methods [Kriegeskorte et al., 2006]. While the methods used for SPMs and MPA approaches are similar, they rely on different assumptions about the data generating process. Some assumptions however the two approaches have in common. For instance all supervised settings, when applied to fMRI data, need to correct for the hemodynamic lag when correlating fMRI signals with a stimulus. This is usually done by convolving the stimulus regressor with a canonical hemodynamic response function (HRF). An example of a typical HRF as implemented in the analysis software SPM [Friston and Buchel, 2003] is shown in figure 4.1. It is based on a biomechanical model of the hemodynamic response, see [Buxton et al., 1998, Friston et al., 2000] and section 4.3. The HRF was generated using the MATLAB function `spm_hrf.m` in SPM5[1]. This HRF is commonly assumed to be the same for all voxels and subjects, hence canonical. While this is a convenient assumption, there is evidence for a considerable variability of HRFs across subjects and brain regions [Aguirre et al., 1998, Handwerker et al., 2004] – neglecting this variability will lead to poor sensitivity to hemodynamic activation which does not follow the canonical HRF dynamics. Chapter 8 will show an alternative approach without these assumptions.

A hemodynamic response function (HRF) models the temporal dynamics of the fMRI signal in response to a neural stimulus

[1]The parameters used were the default parameters, delay of response (relative to onset) 6s, delay of undershoot (relative to onset) 16s, dispersion of response 1s, dispersion of undershoot 1s, ratio of response to undershoot 6s, onset (seconds) 0s, 32s, temporal resolution 1Hz

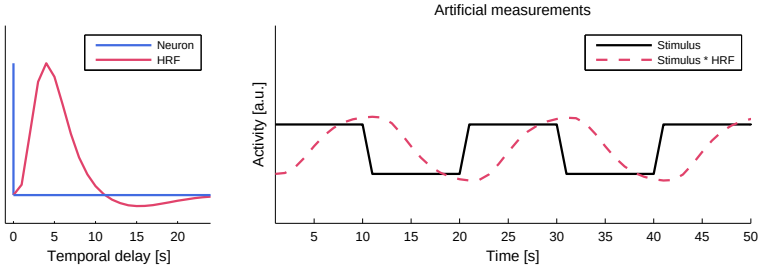

Figure 4.1: Example of a canonical hemodynamic response function (HRF); in order to construct a regressor for the GLM analysis, a stimulus time series (*black*, left panel) is convolved ($*$) with the canonical HRF (*grey*, right panel); the resulting hypothesized fMRI time series is a smoothed version of the stimulus;

Mass-univariate methods In mass-univariate methods one treats single voxel or dipole time series as target variable y and predicts their time course from a linear combination of multivariate regressors x containing experimentally controlled variables and parameters that are not of interest but account for variance y. If the time course is explained well by the regressor of interest, that voxel will have a high weight in the SPM. Most SPM analyses are based on the so called *general linear model* (GLM). GLMs are the most often used class of supervised methods for finding statistical parametric maps (SPMs) of neural activation [Friston et al., 1994, Friston and Buchel, 2003]. For a given data set the target variable $y \in \mathbb{R}^L$ (e.g. a single voxel time course of length L) is modeled as a linear combination of all N regressors, each weighted by a coefficient stored in a vector $\beta \in \mathbb{R}^N$, plus some gaussian i.i.d. error $\varepsilon \sim \mathcal{N}(0,1)$

Statistical parametric maps (SPMs) visualize activation patterns that are correlated with an experimental stimulus

$$\hat{y} = X\beta + \varepsilon. \tag{4.3}$$

The $L \times N$ matrix $X = [x_1\ x_2\ \ldots\ x_N]$ containing the time series of N regressors x_i of length L as column vectors is called *design matrix*. A typical GLM analysis includes as regressors all experimentally controlled parameters and additionally so called *nuisance* regressors, which are not of interest in the analysis but explain some of the variance in the data. In fMRI data, such a nuisance regressor could be for instance head movement within the scanner. Including an estimate of movement along each axis (pitch, roll, yaw) into the design matrix will improve the fit of the GLM. In SPM analyses for fMRI data y is a single voxel time course and the same design matrix X is fitted to all voxels separately. This approach is called *mass-univariate* analysis. Equation (4.3) can be solved by the *ordinary least squares* (OLS) solution

General linear models (GLMs) are the most often used method to compute SPMs

$$\hat{\beta} = (X^\top X)^{-1} X^\top y^\top. \tag{4.4}$$

The magnitude of the entries in the vector β can now be subjected to statistical tests. Voxels with values of β that are significantly different from zero will be considered

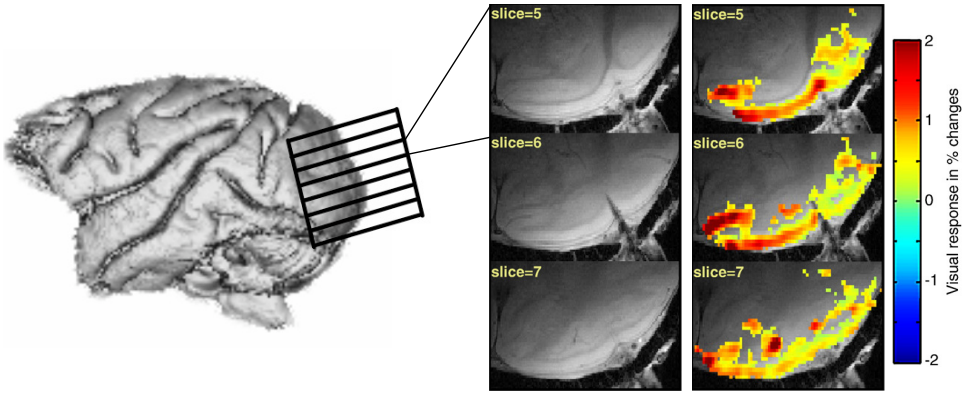

Figure 4.2: Statistical parametric map (SPM) of early visual cortex in the macaque monkey, for experimental details see appendix B or [Murayama et al., 2010]; voxels with significantly non-zero β values are superimposed on anatomical MRI scans; color code corresponds to percent visual activation (relative to baseline);

as 'activated' by the experimental paradigm modeled in the design matrix X. This involves a statistical test for every voxel. Since typically fMRI data contains a few ten thousand voxels it is almost certain that a large fraction of voxels will be considered 'active' due to random variations and not because those voxels are really 'activated'. The statistical threshold for significance has to be adapted to this *multiple comparison problem*. Figure 4.2 shows an example of an activation pattern visualized with an SPM superimposed on anatomical MRI scans of primary visual cortex of a Macaque monkey [Murayama et al., 2010]. Experimental details can be found in appendix B. The visual stimulus shown was a rotating checkerboard that was on for 6 seconds and then off for 24 seconds. This was repeated 10 times. The regressor in the design matrix X (here, $X \in \mathbb{R}^{L\times 2}$, x_1 is the visual stimulus time course and x_2 an offset term) thus was a boxcar function which was 1 if the stimulus was on and 0 otherwise (that is before the convolution with the canonical HRF). This approach has been and continues to be the most popular analysis method for fMRI data. Due to the assumptions underlying SPM estimation there are a number of caveats when dealing with GLM type analyses, for a critical review see e.g. [Monti, 2011]. The most often criticized assumptions are temporal and spatial independence: Each voxel and time sample is treated as an independent variable. These assumptions are violated in real fMRI data. The dependencies present already e.g. in fMRI measurements are even enhanced by standard preprocessing procedures such as temporal smoothing of regressors with a canonical HRF and spatial smoothing of the fMRI data. Temporal smoothing is important to temporally align stimulus time course and voxel time course. Spatial smoothing is needed to increase the signal to noise ratio of fMRI data, to reduce numerical instabilities when co-registering single subject data in a common template brain and to reduce the effective degrees of freedom of the

Real neuroimaging data violate the assumptions of SPM approaches

fMRI data for multiple comparison corrections. Temporal correlations are usually removed from the regressors in the design matrix prior to the GLM fitting procedure [Bullmore et al., 1996].

Multivariate methods As mass-univariate approaches consider only one voxel at a time, certain multivariate voxel patterns cannot be detected [Lange et al., 1999]; in general mass-univariate measures are less sensitive than multivariate approaches [Kriegeskorte et al., 2006]. Consequently there is increased interest among neuroscientists in multivariate methods, often referred to as *multivariate pattern analysis* (MPA)[2]; in the context of fMRI analysis this approach is also known as *multi-voxel pattern analysis* [Norman et al., 2006], *brain-reading* [Cox and Savoy, 2003], *distributed pattern analysis* [Haxby et al., 2001], *mind-reading* [Haynes et al., 2007], *information-based functional brain mapping* [Kriegeskorte et al., 2006] or *prediction* [Haynes and Rees, 2005]. The goal of MPA is to find patterns in the data that predict the stimulus well. A common multivariate approach for fMRI analyses is to predict some stimulus label time course $y \in \mathbb{R}^1$ using a *searchlight*, that means to extract small voxel volumes $v \in \mathbb{R}^U$ (U denotes the number of voxels of the vectorized volume), which are mapped through some function $f(.)$

Multivariate Pattern Analysis (MPA) predicts univariate stimulus labels from multivariate measurements, like voxel volumes of fMRI signals

$$f(v) = \hat{y} + \varepsilon, \tag{4.5}$$

where ε again denotes noise variance that is not explained by $f(v)$. There are plenty of possibilities to estimate $f(.)$ from a given data set of size L of stimulus target variables $y \in \mathbb{R}^{1\times L}$ and voxel time series $V = [v_1, v_2, \ldots, v_L] \in \mathbb{R}^{U\times L}$. Popular examples include *support vector machines* (SVMs) [Schölkopf and Smola, 2001, Müller et al., 2001], relevance vector machines [Tipping, 2001], gaussian processes (GPs) [Rasmussen and Williams, 2005], linear regression and Fisher Linear Discriminant analysis [Fisher, 1936]. For a comparison between some of these algorithms applied to fMRI data see e.g. [Lange et al., 1999]. Although choosing $f(.)$ to be a non-linear function can yield better prediction accuracy, linear methods – where $f(V) := w^\top V, w \in \mathbb{R}^{U\times 1}$ – are more popular [Müller et al., 2003]. First they are easy to compute; secondly there is a straightforward interpretation[3] of the weight vector w: It contains the contribution of each voxel to the prediction $\hat{y} \in \mathbb{R}^{1\times L}$. In the case of linear regression we find w in complete analogy to eq. (4.3), only that X is now the data and the target y is the time series of stimulus labels. When estimating w one is not interested in *learning the data by heart*. Any stimulus time course can be

[2]Of course mass-univariate methods are in principle also multivariate – but only with respect to the experimental variables and not with respect to fMRI voxels; multivariate in the context of fMRI analyses often refers to the notion of a multivariate measurement, as opposed to the mass-univariate treatment of single voxels.

[3]In order to have an interpretable solution w, the data in V should have been normalized, such that each row has the same variance.

predicted at 100% accuracy if the data has enough dimensions. Instead one is interested in a low *generalization error*; this means that w should be such that *new* data, which has not been seen by the learning algorithm, is predicted at high accuracy. This can be achieved by constraining the complexity of the solution w and is called *regularization* [Duda et al., 2001]. In the case of linear regression this can be implemented by requiring $\|w\|_2^2$ to be small. The objective of regularized linear regression is, in analogy to eq. (4.4)

$$\underset{w}{\operatorname{argmin}} \left((y - w^\top X)^2 + \kappa \|w\|_2^2 \right) \tag{4.6}$$

where κ is called *regularization parameter*, controlling the amount of regularization. The larger κ the more important is the 'simplicity' of the solution in the optimization of eq. (4.6). In the worst case of very high noise κ will we chosen to be very large and consequently all entries of w will be small and similar. If all entries of w are the same, as in the case of high noise and hence high regularization, the solution is the same (up to a constant scaling factor) as the mean of X, in which each dimension is weighted with $1/L$. Solving eq. (4.6) for w yields, in analogy to eq. (4.4)

$$w = (X^\top X + I\kappa)^{-1} X^\top y. \tag{4.7}$$

A linear kernel is the inner product of a data matrix and contains a similarity measure for all pairs of data points

The matrix product $X^\top X$ is also called *linear kernel* of X. As the regularization adds a ridge of height κ on this matrix, eq. (4.7) is often referred to as *kernel ridge regression*. In order to find the correct κ, classical statistical learning routines such as *cross-validation* can be used, see e.g. [Lemm et al., 2011].

4.2 Unsupervised Methods

Not in all experimental settings a regressor or target variable, needed for supervised analyses, is available. And often there are many more processes reflected in the signal than the experimentally controlled stimulation. Many of these factors cannot be measured or modeled. Given the same stimulus, repeated measurements of brain activity will show variations around the mean response to that stimulus. These variations might be considered as noise as they do not reflect the stimulus - but just because we cannot explain certain aspects of neural activity by a stimulus (or additional nuisance regressors) that does not mean that this variability is noise. It merely means that we do not have a model for it. If there is no model for a measurement, *unsupervised methods* can be used to find structure in the data in an explorative manner. This can help to find interesting patterns in the data, for instance interesting in the sense that a pattern found by explorative methods coincides with anatomical findings obtained independently. In practice most unsupervised projection methods can be cast in the form of

$$Y = WX \tag{4.8}$$

where $X \in \mathbb{R}^{M \times L}$ is the data matrix, $W = [w_1^\top, w_2^\top, \ldots, w_n^\top]^\top \in \mathbb{R}^{N \times M}$ is a linear mapping that transforms the data in X into a new representation $Y \in \mathbb{R}^{N \times L}$. Each basis vector w_i computes a linear combination of the M original dimensions of X. How the original dimensions of X are combined depends on the criterion optimized by the respective unsupervised learning method used. In the following we will shortly introduce two popular examples of unsupervised learning techniques, *principal component analysis* (PCA) and *independent component analysis* (ICA). Both have become standard techniques in the context of neuroimaging.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) finds directions of maximal variance in the data

PCA [Pearson, 1901] finds a linear combination of features (such as sensors, voxels, etc.) that maximizes the variance of the data. Inspecting the feature combinations gives a good estimate of which combination of sensors captures most variance of the signal. And projecting the data onto those feature combinations allows to reduce a high dimensional data set to only a few dimensions that can be easily visualized. PCA is probably the most often used method for dimensionality reduction. Consider a data set stored in a matrix $X \in \mathbb{R}^{D \times L}$ (D dimensions and L data points). PCA computes a linear projection $w^* \in \mathbb{R}^D$ that maximizes the variance in the data set:

$$w^* = \underset{w}{\operatorname{argmax}}\, w^\top X X^\top w \quad \text{s.t. } \|w\| = \sqrt{w w^\top} = 1. \tag{4.9}$$

The matrix product $lXX^\top, l = 1/L$ is called *covariance matrix* C of X (we will omit the normalization constant l in the following). It contains in its ith row and jth column the covariance between the ith and jth dimension of X. We will introduce two ways of solving PCA, first the standard solution and another version that does not require the explicit computation of covariance matrices; this is advantageous in case the data is high-dimensional. There are probabilistic versions of PCA in which the underlying generative model is a multivariate gaussian distribution [Roweis and Ghahramani, 1999]. The principal components $W = [w_1, w_2, \ldots, w_D]$ (we assume here that $D \leq L$) are the eigenvectors of the data covariance matrix C

$$CW = W\Lambda, \tag{4.10}$$

Principal components are eigenvectors of the data covariance matrix

where the matrix Λ contains on its diagonal the corresponding eigenvalues $\lambda_1, \ldots, \lambda_k$ (and is 0 in its off-diagonal entries). As Λ is the new covariance matrix of the data in the PCA space, PCA is often referred to as *diagonalization* of the covariance matrix. The basis vectors of W are aligned with the directions of maximal variance in the data, as depicted in fig. 4.3. Looking at the eigenvalues helps to decide how many principal components should be considered when analyzing the data. A fast decay after i eigenvalues is a good sign that the data is well captured by the first i principal components. We can transform the data X into its new representation X_{PCA} in

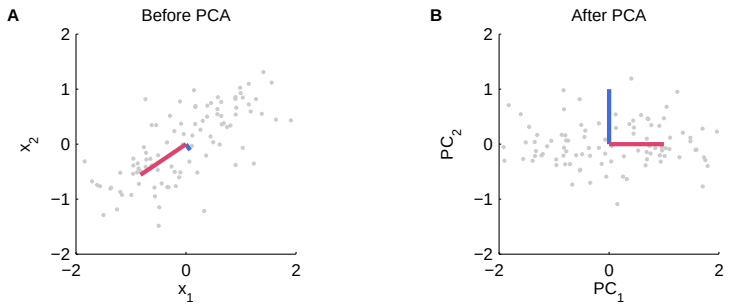

Figure 4.3: Two dimensional toy data example for PCA; in all panels, the first dimension is plotted on the x-axis, the second dimension along the y-axis; **A**: data was drawn from a two dimensional gaussian distribution; the first and second principal component directions are plotted in green and red, respectively; **B**: The same data, projected onto the first two principal components; the direction of the data that exhibits most variance is now aligned with the x-axis and the direction of the second largest variance is parallel to the y-axis;

which all rows of the data matrix X are uncorrelated

$$X_{PCA} = W^\top X = \sum_{i=1}^{k} e_i w_i^\top X, \tag{4.11}$$

PCA can be used to discard low variance directions (for dimensionality reduction) or high variance directions (for artifact removal)

where the e_i are the basis vectors of the canonical euclidean basis. An illustration with two dimensional toy data is shown in figure 4.3. The measurements x (in grey) are drawn from a two dimensional gaussian distribution. The directions of maximal variance, the principal components, are plotted on top of the data in fig. 4.3A. After projecting the data onto their first two principal components, the axes of maximal variances of X_{PCA} are rotated such that they fall on the main coordinates (fig. 4.3B). Often PCA is used for dimensionality reduction. In this case one is interested in a lower dimensional representation of the data, capturing as much variance as possible. This can easily be achieved by retaining only the first rows of X_{PCA}, i.e. those that correspond to the highest eigenvalues. Since W is orthogonal we have $X = WW^\top X$ and can therefore remove certain components also in the original basis of the data by

$$X' = \sum_{i \in \mathcal{I}} w_i w_i^\top X \tag{4.12}$$

where the index set $\mathcal{I}$ contains all component indices that should be retained. Even though some components might be removed, the new data set X' is expressed in the same basis as X and therefore has the same number of rows.

PCA for high-dimensional data If the dimensionality D of the data is higher than the number of samples L it becomes more efficient to not work on covariance ma-

trices but on *inner products* of the data, also known as the *linear kernel* matrix K_X, see also eq. (4.7), of the data

$$K_X = X^\top X. \tag{4.13}$$

The linear kernel K_X of the data is only of size $L \times L$, while the covariance matrix would be of size $D \times D$. We will consider here only linear kernels. An advantage of *linear* kernels is that the projections w can be easily recovered as a linear combination of data points. For the first principal component w this means:

$$w = X\alpha \tag{4.14}$$

Where $\alpha \in \mathbb{R}^L$ is a column vector containing the contribution of each data point in X to principal component w. The intuition behind this expansion is that if the number of dimensions is larger than the number of data points, then the solution has to lie in the space spanned by the data. Thus we can express the solution w as a combination of data points, each weighted according to its respective coefficient in α. The relationship between linear kernel PCA and standard covariance matrix based PCA as computed in eq. (4.10) is easy to understand when considering the singular value decomposition of X, see appendix A.2. If each data point contributes exactly the same w is the mean over all data points. Next to the more efficient computation (in some settings) compared to diagonalizing the covariance matrix, the kernel version of PCA has another advantage: Non-linear extensions are straightforward to implement [Müller et al., 2001, Schölkopf and Smola, 2001] by computing the covariance not in the input space but in a high-dimensional feature space defined by non-linear *kernel functions* [Schölkopf et al., 1998]. This *kernel-trick* can be helpful if the structure of the data is not well captured by a linear feature combination. Chapter 7 will explain the kernel-trick in more detail.

Independent Component Analysis (ICA)

Independent Component Analysis (ICA) is another unsupervised algorithm that maximizes statistical independence between the components, i.e. the rows of matrix Y in eq. (4.8). The ICA model can be formulated as

$$X = AS, \tag{4.15}$$

meaning that the nth observed time series $x_n \in \mathbb{R}^{1\times L}$ (L denoting the number of time points sampled) is a linear and instantaneous mixture of M statistically independent *source signals* $s_m \in \mathbb{R}^{1\times L}$ with $m = 1, \ldots, M$. The matrix A is usually referred to as *mixing matrix*. The goal of ICA is to reconstruct the original source signals S given only the mixtures X. Maximizing the independence of the source estimates $\hat{S} \in \mathbb{R}^{N\times L}$ corresponds to minimizing the *mutual information* (see section

5.3) between the estimated source signals

$$I(\mathcal{S}) = \sum_n H(\mathcal{S}_n) - H(\mathcal{S}), \tag{4.16}$$

where $\mathcal{S}$ denotes the random variable of the estimated source signals $\hat{S}$, $H(\mathcal{S}_n)$ denotes the *entropy* of the single source estimates (the nth row of $\hat{S}$) and $H(\mathcal{S})$ the total entropy of the source estimates. Over the last 15 years, a great number of different algorithms have been proposed to find these non-gaussian projections; some optimize contrast functions like the negentropy by gradient descent [Bell and Sejnowski, 1995], others solve the ICA problem by approximate diagonalization of the fourth-order cumulant tensor (kurtosis) [Cardoso and Souloumiac, 1993], by fixed-point algorithms [Hyvärinen, 1999] or by mutual information based approaches employing kernel canonical correlation analysis [Bach and Jordan, 2002]. Many ICA software packages have been developed and enable non-experts to use the most popular ICA implementations on EEG/MEG and fMRI data.

Spatial and Temporal ICA Each column of the mixing matrix A in eq. (4.15) reflects one source signal and can be interpreted as the *field pattern* that shows how this source contributes to the measured activity in all the different channels (EEG/MEG-channels or fMRI-voxels). According to this interpretation the right side of equation (4.15) is the sum of outer products between space vectors A and time vectors S, where the time vectors are independent. Therefore, this setting is often called *temporal ICA*. Another variant is called *spatial ICA*. Here the roles between row- and column vectors have switched, i.e. instead of independent time series one computes independent spatial patterns. This corresponds to performing ICA on $X^\top$ instead of X. Spatial ICA on fMRI data is often used for *functional connectivity analysis* [van de Ven et al., 2004, Ystad et al., 2011]. For temporal ICA, the number of time points should be larger than the number of channels (i.e. $T > N$), for spatial ICA the number of time points should be smaller than the number of dimensions ($T < N$). Therefore the former is suited for typical EEG and MEG analysis, while the latter is the method of choice for fMRI data. For a more complete discussion about differences between spatial and temporal ICA and the assumptions underlying applications of ICA on fMRI data see e.g. [McKeown and Sejnowski, 1998, Calhoun et al., 2003]. For a review on ICA for fMRI data analysis see e.g. [Calhoun et al., 2009].

4.3 Model Based Approaches

Most of the above mentioned analysis methods can be regarded as data driven. In the majority of studies the only model used is the canonical HRF. A problem with data driven analysis methods is that they do not offer a physiologically interpretable

model of the underlying neural processes. However, if one aims at testing hypotheses about physiological parameters, a well defined physiological model of the signal of interest is indispensable. Hypotheses, or prior knowledge about anatomical and functional properties, can be incorporated in models of the physiological processes underlying the measurements. These physiological models are called *forward models*. A forward model $\mathcal{F}$ is specified by a set of equations describing the dynamics of electromagnetic or hemodynamic activity and parameters φ. In analogy to the standard supervised learning setting , see e.g. eq. (4.2), these parameters are optimized such that some distance measure (e.g. the euclidean distance) between the measured data $\mathcal{D}$ and the model predictions $\mathcal{F}(\varphi)$ is minimized.

$$\underset{\varphi}{\operatorname{argmin}} \left(\|\mathcal{D} - \mathcal{F}(\varphi)\|_2 \right). \tag{4.17}$$

The following two paragraphs section give two examples of popular forward models for EEG and fMRI data. A more detailed overview of model driven analysis approaches in the context of multimodal data can be found in [Bießmann et al., 2011b].

EEG/MEG forward models

The physical quantities modeled in EEG/MEG forward models are current dipoles between a current source and sink. A dipole is a convenient representation of the electrical field emerging from synchronized activation of a neural cell assembly. The strength of a single electrical dipole moment $s \in \mathbb{R}^3$ (three orthogonal spatial components) is related to electrode potentials measured in $x \in \mathbb{R}^U$ (U denotes number of electrodes) by the forward model

$$x = As \tag{4.18}$$

where $A \in \mathbb{R}^{U \times 3}$ contains in its rows the *lead-field* of the respective electrode, reflecting how much of the dipole activity is picked up by that electrode; in its columns A contains the *gain* of each electrode. This matrix models the conductivity profile of the head and can be estimated using anatomical measurements obtained by structural MRI [Dale et al., 2000]. The linear relationship in eq. (4.18) relies on the assumption that magnetic induction and capacitive effects do not influence the measurements. Adding N dipoles results in a dipole vector $s \in \mathbb{R}^{3N}$ and a lead-field matrix $A \in \mathbb{R}^{U \times 3N}$. The forward model in A can be estimated using e.g. boundary element methods [Oostendorp and van Oosterom, 1989]. In order to recover the source activity the forward model has to be inverted

$$\hat{s} = Wx, \tag{4.19}$$

where $\hat{s}$ denotes estimated source time courses and $W \in \mathbb{R}^{3N \times U}$ is the inverse solution that maps measurements to sources. Many approaches to computing W have

been proposed, including a simple least squares approach as in section 4.1, bayesian model inversion [Sato et al., 2004] or ICA based techniques (see section 4.2), to name just a few. A popular solution to the model inversion problem is called *minimum norm* approach [Dale and Sereno, 1993]

$$W = RA^\top (ARA^\top + \kappa E)^{-1} \tag{4.20}$$

where $E \in \mathbb{R}^{U\times U}$ is the sensor noise covariance matrix, κ is the regularization parameter and $R \in \mathbb{R}^{3N\times 3N}$ is the dipole source covariance matrix. In fMRI constrained current source density analysis [Liu et al., 1998, Dale et al., 2000, Liu and He, 2008] R is assumed to be a diagonal matrix which contains on its diagonal spatial weighting factors obtained from anatomical MRI scans or from GLM type analyses (see section 4.1). For a comprehensive review of the physiological origins of EEG/MEG see [da Silva and Niedermayer, 1998, Nunez and Srinivasan, 2006]. A detailed review on solutions to EEG/MEG forward models and model inversion can be found in [Baillet, 2001]. For realistic head models there are no analytical solutions and numerical solutions for realistic head models are computationally costly.

Forward Models for fMRI data

A biomechanical model of the BOLD signal is proposed in [Buxton et al., 1998]. This so called *balloon model* was extended by [Friston et al., 2000] to reflect the complete process of the BOLD response in a veneous balloon starting from synaptic activity $u(t)$. For each location in the brain the temporal dynamics of four physiological quantities are modeled, a flow inducing signal ξ, blood inflow φ, venous volume v and deoxyhemoglobin concentration ω. The model is based on six parameters, neuronal efficacy ε, signal decay τ_s, auto-regulation τ_f, transit time τ_0, stiffness parameter α and resting oxygen extraction E_0. The differential equations of the balloon model are:

$$\dot{\xi} = \varepsilon u(t) - \frac{\xi}{\tau_s} - \frac{\varphi - 1}{\tau_f} \tag{4.21a}$$

$$\dot{\varphi} = \xi \tag{4.21b}$$

$$\dot{v} = \frac{1}{\tau_0}(\varphi - f_{out}(v,\alpha)) \tag{4.21c}$$

$$\dot{\omega} = \frac{1}{\tau_0}\left(\varphi\frac{E(\varphi,E_0)}{E_0} - f_{out}(v,\alpha)\frac{\omega}{x_3}\right), \tag{4.21d}$$

where

$$f_{out}(v,\alpha) = v^{1/\alpha} \tag{4.22}$$

$$E(\varphi,E_0) = 1 - (1-E_0)^{1/\varphi} \tag{4.23}$$

The quantity f_{out} relates the flow-volume relationship through the exponent α, which has been determined in experiments to be in the range of 0.1-0.4 [Friston et al., 2000]. Equation (4.23) describes the fraction of oxygen extracted from the inflowing blood. Based on these equations, the authors of [Buxton et al., 2004] propose a canonical hemodynamic response function as depicted in fig. 4.1. Similar to the case of electrophysiological forward models, there is no analytic solution to eq. (4.21). Hence model inversion requires numerical solutions, which are computationally demanding for realistic data set sizes and numerically instable, given the low temporal resolution of fMRI data. It is also important to note that this system of differential equations neglects spatial dependencies. Often studies model only the time course of the average BOLD signal within a given ROI as there is no gold standard model available for spatial and spatiotemporal dynamics. Other studies employ efficient implementations of probabilistic forward models to tackle full brain hemodynamic models, see e.g. [Riera et al., 2007, Daunizeau et al., 2007, Valdes-Sosa et al., 2009]. But some of the efficiency comes at the price of assumptions which are violated in real fMRI data. An example is the spatiotemporal separability assumption of the hemodynamic response function in [Riera et al., 2007, Daunizeau et al., 2007].

Part II

Multimodal Neuroimaging

Chapter 5

Multimodal neuroimaging

In the past decade multimodal neuroimaging methods have become an indispensable tool for neuroscientific research and clinical application [Friston, 2009]. Multimodal neuroimaging methods combine single imaging modalities that yield complementary views on brain activity, such as electrophysiological and hemodynamic measurements (see figure 5.1). Multimodal setups offer many advantages compared to unimodal measurements. In clinical applications for instance one needs high temporal resolution to capture the temporal dynamics of (non-convulsive) epileptic activity and at the same time high spatial resolution to determine the origin of the seizure in the brain. Here, multimodal setups can help to overcome limitations of classical unimodal setups [Vulliemoz et al., 2009, Rosenkranz and Lemieux, 2010]: While unimodal methods typically do not have both high temporal *and* high spatial resolution, multimodal methods combine the advantages of single modalities yielding a view on brain activity with an unprecedented spatiotemporal resolution. Secondly, multimodal methods can enhance our understanding of how neural activity is reflected in single imaging modalities. While fMRI has been established already more than two decades ago [Ogawa et al., 1990] it was not before multimodal methods were established that one could actually measure how neural signals are related to the fMRI signal [Logothetis et al., 2001, Devor et al., 2003]. Knowing how neural activity is reflected in different modalities is important for a better interpretation of the measurements. But there are questions beyond these epistemological issues for which multimodal neuroimaging is indispensable. A better understanding of the mechanisms underlying *neurovascular coupling* is of paramount importance for deeper insights in the pathology of e.g. epilepsy [Vulliemoz et al., 2010a], stroke and Alzheimer's disease [Iadecola, 2004, Girouard and Iadecola, 2006]. Therefore the investigation of neurovascular coupling mechanisms is one of the most important applications of multimodal imaging. This chapter will first give a short overview about multimodal neuroimaging setups in section 5.1, followed in section 5.2 by motivating application examples of these setups in a clinical as well as a research context. Finally section 5.3 will provide an information theoretic view on multimodal data fusion. The information theoretic view can be helpful to understand the different approaches to multimodal data fusion in later chapters.

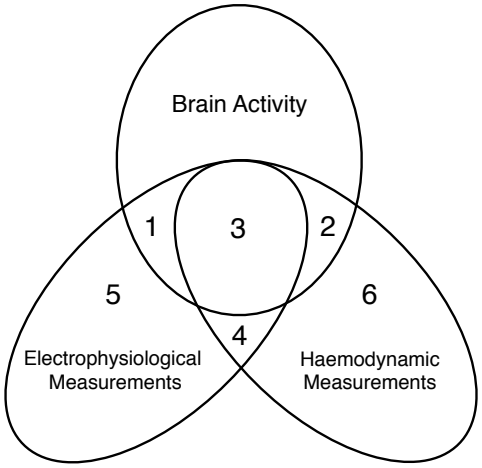

Figure 5.1: Venn Diagram of multimodal neuroimaging analysis methods; some aspects of the brain activity are reflected in electrophysiological recordings and others in hemodynamic measurements; certain aspects such as fast neuronal oscillations are *only* detectable in electrophysiological signals (area **1**), others (such as activity in deep brain structures) are better visible in fMRI signals (area **2**); aspects that are reflected in both modalities can be subdivided into signals originating from neural activity (area **3**) and non-neural physiological processes reflected in both modalities, such as muscle contractions that lead to head movement (area **4**); besides these common artifact sources, there are many artifacts that are reflected in one modality only (area **5** and **6**).

5.1 Multimodal imaging setups

Many different multimodal neuroimaging setups have been established. This section will give a short introduction to the most popular multimodal setups that combine hemodynamic and electrophysiological measurements of brain activity.

Invasive electrophysiology and optical imaging

Combined optical imaging and electrophysiology setups [Grinvald et al., 1986] are comparatively simple and have a high spatiotemporal resolution on the cortical surface. These setups are often used in animal studies for investigation of neurovascular coupling phenomena, see e.g. [Martindale et al., 2003, Devor et al., 2005, Berwick et al., 2008, Sirotin et al., 2009]. Also for human patients the intrinsic optical BOLD signal is an attractive diagnostic tool for clinical applications [Arthur and Nader, 2002]. A disadvantage is that the optical signal is restricted to the cortical surface, while microelectrode measurements can be obtained from deep brain structures.

Setup	Properties	Invasive
Optical imaging-Microelectrodes [Grinvald et al., 1986]	High spatiotemporal resolution **Optical signals restricted to cortical surface**	Yes
fMRI-MEG [Dale et al., 2000]	MEG restricted to cortical surface **No simultaneous measurements**	No
fMRI-EEG [Lemieux et al., 2001]	EEG restricted to cortical surface **Simple Setup**	No
fMRI-Microelectrodes [Logothetis et al., 2001]	**Depth resolution in both modalities** Complex setup	Yes
NIRS-EEG [Moosmann et al., 2003]	**Both modalities restricted to cortical surface** Portable Setup	No

Table 5.1: Popular multimodal neuroimaging setups;

fMRI and MEG

Combined fMRI-MEG measurements were amongst the first multimodal setups. A fundamental limitation of fMRI-MEG is that measurements cannot be obtained simultaneously. In order to analyze MEG-fMRI jointly an experimental paradigm is needed that triggers the same neural activation patterns. Compared to EEG-fMRI setups, the combination of MEG-fMRI might offer the advantage that MEG signals often have a better spatial resolution than EEG setups. In contrast to simultaneous EEG-fMRI, sequential MEG-fMRI measurements have no scanner induced artifacts.

fMRI and EEG

In contrast to fMRI-MEG measurements fMRI-EEG can be measured simultaneously. The first studies using EEG-fMRI had to solve a number of technical challenges [Lemieux et al., 1997, Allen et al., 1998, Goldman et al., 2000, Allen et al., 2000, Lemieux et al., 2001]. After these problems were solved, fMRI-EEG has become the probably most often used multimodal neuroimaging setup. With the advent of commercially available MRI compatible EEG setups and efficient artifact removal methods, this methodology has become a standard tool in clinical applications and neuroscientific research. An fundamental limitation of fMRI-EEG is that the EEG measurements are limited to the cortical surface. Also, the EEG sensors pick up only superpositions of multiple neural dipoles, due to the volume conduction effects of the skull.

NIRS and EEG

The simple and comparatively low-cost setup of NIRS allows to measure hemodynamic activity in many situations when fMRI measurements are not feasible, e.g. for longterm monitoring at the bedside [Bozkurt et al., 2005] or even outside the lab via

wireless transmission [Muehlemann et al., 2008]. For multimodal measurements NIRS has the additional advantage that the electrophysiological measurements are not corrupted by artifacts induced by changes in the strong magnetic fields needed for measuring fMRI signals. Besides NIRS can be measured at temporal resolutions of 25Hz and more, while fMRI measurements usually have temporal resolutions below 1Hz. Of course the temporal response properties of the hemodynamic signal do not change by faster measurements so the effective temporal resolution of the measured hemodynamic response is the same as that of fMRI measurements, but higher temporal resolution can help to understand better the temporal dynamics of the blood response to neural activation. Taken together despite the low spatial resolution and SNR of NIRS, it is an attractive alternative to fMRI in particular for multimodal measurements in everyday life at the bedside of patients. A promising example are applications in the context of brain-computer interfaces (BCIs). In BCIs a computer program extracts volitionally controlled brain signals in order to allow paraplegic patients to communicate with the outside world. So called *hybrid BCIs* [Pfurtscheller et al., 2010], BCIs that use multimodal measurements, can robustly increase the accuracy at which brain states can be decoded from brain activity measurements. For a hybrid NIRS-EEG BCI see for instance [Fazli et al., 2011].

fMRI and intracranial microelectrode recordings

The combination of fMRI and intracranial microelectrode recordings offers the best spatiotemporal resolution of all multimodal setups [Logothetis et al., 2001, Goense and Logothetis, 2008]. All other setups, non-invasive multimodal setups like fMRI-MEG/EEG as well as optical imaging signals combined with intracranial recordings, have at least one modality that can only measure brain activity in the cortical mantle. Intracranial microelectrode recordings combined with high resolution fMRI can measure brain activity in deep structures such as the thalamus in both modalities at the same time [Logothetis et al., 2010]. Simultaneous electrophysiology and fMRI recordings are technically challenging and require specialized equipment [Oeltermann et al., 2007]. A major problem are the artifacts in the electrophysiological recordings induced by switching magnetic field gradients. In the initial part of the work on this thesis a method to remove these artifacts online during a running multimodal measurement was developed and is presented in section 8.5.

5.2 Motivations for multimodal neuroimaging

Multimodal neuroimaging offers decisive advantages for basic research and clinical applications. A few motivating examples will illustrate the usefulness of multimodal methods in the following.

Understanding unimodal concepts of brain activity

Most of what we know about brain activity in healthy human subjects has been found with unimodal measurements of neural activity using EEG, MEG, fMRI, PET and invasive microelectrode recordings. Consequently most neuroscientific concepts are tightly related to the method that has been used to discover and investigate this particular aspect of neural activity. Translating unimodal concepts from one modality to another is an important step towards understanding the neural mechanisms underlying these phenomena.

From EEG concepts to fMRI An important concept in the EEG literature are event related responses (ERPs) and amplitude modulations of neural oscillations. Both, the amplitude of oscillations in a certain frequency band as well as ERPs are being used as diagnostic markers in clinical practice. ERPs are computed by taking the average of EEG/MEG epochs time locked to the onset of a stimulus event. Systematic peaks and troughs in the EEG/MEG signal after stimulus onset will add up in the average and can be used to infer the temporal dynamics of cortical processing. A better understanding of the neural mechanisms underlying the generation of ERPs requires a multimodal approach. In [Eichele et al., 2005] multimodal EEG-fMRI is used to link ERPs to their fMRI representation. In an auditory target detection experiment, the amplitude modulation at different latencies was used to predict spatially separated fMRI activation patterns, reflecting different stages in the hierarchy of cortical processing. Besides ERPs a well studied EEG feature is the average amplitude of neural oscillations within a certain frequency band. The best studied oscillation is the so called α rhythm[1] [Berger, 1929]. Concurrent EEG-fMRI studies could relate the modulation of alpha power to decreased BOLD signals in occipital cortex and increased BOLD signals in thalamus [Goldman et al., 2002, Moosmann et al., 2003, de Munck et al., 2008, Wu et al., 2010]. Oscillations in other frequency bands have been studied with EEG-fMRI in [Laufs et al., 2003, Scheeringa et al., 2008].

From fMRI concepts to EEG An example of a unimodal concept from fMRI data are so called *resting state networks* [Laufs et al., 2003, van de Ven et al., 2004, Cole et al., 2010]. These networks are estimated by analyzing ongoing hemodynamic activity in the absence of any task. Although the functional relevance of these networks is currently not fully understood there is evidence to suggest that they represent a fundamental and functionally relevant modular organization [Smith et al., 2009]. It has been proposed that their integrity is compromised in a variety of neurologic and psychiatric conditions [Greicius et al., 2003, Gotman et al., 2005]. A prominent

[1] The α rhythm is characterized by its posterior distribution and oscillations in a frequency range from about 8 to 12 Hz and reactivity to eye opening/closure, i.e. closing the eyes will result in increased amplitudes in these oscillations, while opening leads to amplitude decrease.

example is the *default mode network* [Raichle et al., 2001]. Per definition the *default mode network* is a fMRI concept as it is based on the *oxygen extraction fraction*, the ratio of oxygen used by the brain divided by the amount of oxygen delivered by the vessels. As the metabolic needs of the brain in response to sensory stimulation are rather subtle compared to its energy consumption during rest, the authors of [Raichle et al., 2001] advocate the idea that it is crucial for our understanding of neural processing to define a *baseline* state of brain activity. Studies investigating the *default mode network* are lacking experimentally controlled variables that could be used as regressor for analyzing the network of interest; hence resting state data requires specialized analysis approaches. Before multimodal neuroimaging methods were established most resting state studies have been relying on unimodal *unsupervised* methods (see section 4.2). As non-invasive multimodal EEG-fMRI setups became readily available, researchers were able to study the these hemodynamic phenomena with supervised methods by using one modality as regressor for the other modality [Laufs et al., 2003, Mantini et al., 2007, de Munck et al., 2008, Laufs, 2008, Ritter et al., 2009].

Diagnosis of pathological brain activity

In clinical research the development of concurrent EEG-fMRI protocols was motivated by the aim to improve seizure localization for pre-surgical planning of pharmacoresistant epilepsy. It is crucial for surgeons to localize epileptic sources as precisely as possible in order to not damage structures that are important for functions such as memory, speech or motor coordination. Moreover it is important to understand the mechanisms underlying abnormal brain activity in these patients, such as the spatial extent of the seizure generating network. Both questions have been addressed in studies employing multimodal methods. Several groups have explored and demonstrated the feasibility of simultaneous intracranial, EEG and fMRI or NIRS recordings for this purpose [Ives et al., 1993, Krakow et al., 1999, Lemieux et al., 2001]. Also intracranially measured interictal[2] EEG events in epileptic patients events have been used as predictors of hemodynamic activity measured in fMRI [Vulliemoz et al., 2011]. The basic assumption in this approach is that hemodynamic responses at the time of the interictal events would map on to the regions of the brain that are involved in generating seizures - the irritative zone(s). This principle was then also applied in studies of patients with generalized non-convulsive seizures in which spike-wave discharges were used to predict hemodynamic responses [Gotman et al., 2005], as well as partial seizures. Other studies investigated the hemodynamic response of epileptic activity in children using simultaneous EEG-NIRS measurements [Roche-Labarbe et al., 2008]. The authors report an increase in deoxyhemoglobin shortly before the onset of EEG spike-and-wave discharges. As NIRS is, like fMRI, a mea-

[2] *Ictal* refers to the period of an epileptic seizure. *Interictal* refers to the period between epileptic seizures during which EEG events characteristic for epileptic disorders can be detected.

sure of metabolic activity, this study shows that epileptic activity measured by electrophysiological recordings is preceded by an increase in metabolic rate. This is in line with previous studies in animal models [Zhao et al., 2007] and hints at the potential of multimodal methods for epilepsy research: Without multimodal measurements it is impossible to find such a *dissociation* between neural and hemodynamic activity. Often the localization of epileptic sources improves with multimodal measurements. The accuracy of multimodal estimates of epileptic sources can be validated with intracranial electrophysiology [Vulliemoz et al., 2011] and quantification of post-surgical seizure reduction [Thornton et al., 2010a]. However there are many patients in whom localization of epileptic sources from one modality does not coincide with that of another modality. Novel analysis approaches might help to solve these discrepancies [Jann et al., 2008, Thornton et al., 2010b, Vulliemoz et al., 2010b].

5.3 An information theoretic view

For an intuitive understanding of multimodal analyses it is helpful to take an information theoretic view. Information theory has been developed as a generic framework for telecommunication [Shannon, 1948] and became popular in application domains far beyond that [Cover and Thomas, 2006]. Independent of the modality, not all aspects of brain activity can be measured; and not for all aspects measured in one modality there might be a counterpart in other modalities. Besides especially in multimodal setups artifacts impose additional challenges on the analyses. Information theoretic measures can help formalizing and illustrating these relationships. The basic quantity underlying information theory is the *entropy* $H(\mathcal{X})$ of a random variable $\mathcal{X}$ with probability distribution $p(x)$; it is measured in bits and defined as

$$H(\mathcal{X}) = -\sum_x p(x) log_2(p(x)). \tag{5.1}$$

Entropy can be used to derive another measure, the *mutual information* $I(X;Y)$ between X and another variable Y, which is the reduction in entropy of X by knowing the state of Y

$$I(\mathcal{X};\mathcal{Y}) = H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y}) \tag{5.2}$$

where the *conditional entropy* $H(\mathcal{X}|\mathcal{Y})$ is given by

$$H(\mathcal{X}|\mathcal{Y}) = -\sum_x \sum_y p(x,y) log_2\left(p(y|x)\right). \tag{5.3}$$

All quantities are illustrated in figure 5.2A. The mutual information $I(\mathcal{X};\mathcal{Y})$ is the intersection of the areas corresponding to $H(\mathcal{X})$ and $H(\mathcal{Y})$, respectively. Taking the information theoretic stance one can easily relate theoretical insights to simple

geometrical intuitions. In particular we see that the mutual information can also be computed by

$$I(\mathcal{X};\mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}) - H(\mathcal{X},\mathcal{Y}) \tag{5.4}$$

which generalizes in the case of more than two variables $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_N$ to

$$I(\mathcal{X}_1;\mathcal{X}_2;\ldots;\mathcal{X}_N) = \sum_{n=1}^{N} H(\mathcal{X}_n) - H(\mathcal{X}_1,\ldots,\mathcal{X}_N). \tag{5.5}$$

This relationship is used for instance in ICA (see section 4.2). We will denote the random variable of all (discrete) brain states $\mathcal{Z}$, the variable describing all EEG measurements $\mathcal{X}$ and that of all fMRI states $\mathcal{Y}$. The joint entropy of all variables $H(\mathcal{X},\mathcal{Y},\mathcal{Z})$ is illustrated in figure 5.1 as the joint area within all three ellipses. The goal of multimodal neuroimaging is to find those features in the measurements that maximize the grey area in figure 5.2B, which is the mutual information

$$I(\mathcal{Z};\mathcal{X},\mathcal{Y}) = H(\mathcal{Z}) - H(\mathcal{Z}|\mathcal{X},\mathcal{Y}) \tag{5.6}$$

between the brain activity $\mathcal{Z}$ and the joint entropy of all measurements $\mathcal{X},\mathcal{Y}$. If an analysis method picks those features that maximize equation (5.6), the entire area of $H(\mathcal{Z})$ intersecting with either $H(\mathcal{X})$ or $H(Y)$ is covered. The remaining area of $H(\mathcal{Z})$ is the conditional entropy $H(\mathcal{Z}|\mathcal{X},\mathcal{Y})$. If there is information gained by multimodal measurements the area covered by $I(\mathcal{Z};\mathcal{X},\mathcal{Y})$ is larger than the areas of $I(\mathcal{Z};\mathcal{X})$ or $I(\mathcal{Z};\mathcal{Y})$, representing unimodal measurements. As the true brain states $\mathcal{Z}$ are unknown, eq. (5.6) cannot be maximized directly. Instead one often maximizes

$$I(\mathcal{S};\mathcal{X},\mathcal{Y}) = H(\mathcal{S}) - H(\mathcal{S}|\mathcal{X},\mathcal{Y}) \tag{5.7}$$

the mutual information between a stimulus $\mathcal{S}$ and the measurements. As the stimulus entropy is fixed by the experimenter, eq. (5.7) reduces to minimizing the entropy $H(\mathcal{S}|\mathcal{X},\mathcal{Y})$; in most analysis settings this translates into *explaining as much of the stimulus variance as possible using the measurements*. Few studies formulate this objective explicitly, but for a comparison of the various analysis approaches the information theoretic intuitions illustrated in figure 5.2 can be helpful.

Other approaches are to maximize $I(\mathcal{X};\mathcal{Y})$, the mutual information between multiple measurements (area 3 and 4 in fig. 5.2). This is equivalent to the aforementioned approaches, but it requires thorough removal of all artifacts (corresponding to area 4) and proper feature extraction covering all relevant parts of areas 1 and 2, prior to maximizing $I(\mathcal{X};\mathcal{Y})$. Examples of this approach are [Ostwald et al., 2010, 2011].

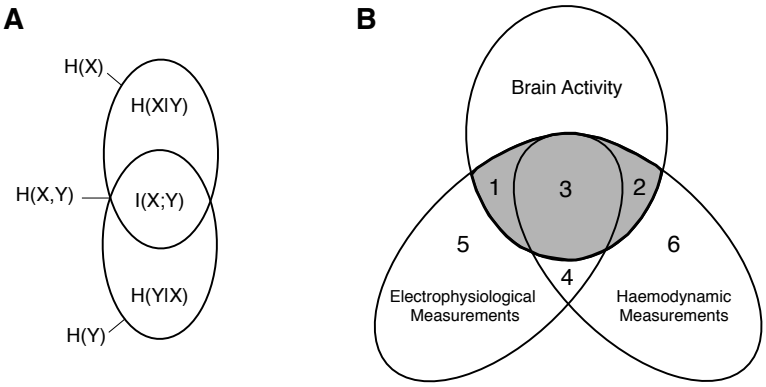

Figure 5.2: Information theoretic illustration of multimodal analyses; **A** Information theoretic quantities visualized as areas in a Venn diagram; **B** Entropies of brain activity labeled as in fig. 5.1; area **1** denotes brain activity that is reflected only in electrophysiological measures, area **2** activity reflected only in hemodynamic signals and area **3** brain activity reflected in both modalities; ideally multimodal analysis methods maximize the mutual information $I(\mathcal{Z};\mathcal{X},\mathcal{Y})$ (see eq. (5.6)) between brain activity $\mathcal{Z}$ and all measurements $\mathcal{X},\mathcal{Y}$ (corresponding to electrophysiological and hemodynamic signals, respectively), comprising areas **1**-**3** (grey); area **1** is that part of $I(\mathcal{Z};\mathcal{X},\mathcal{Y})$ that is not contained in the mutual information between $\mathcal{Z}$ and $\mathcal{Y}$, so area **1**=$I(\mathcal{Z};\mathcal{X},\mathcal{Y}) - I(\mathcal{Z};\mathcal{Y})$, area **2**=$I(\mathcal{Z};\mathcal{X},\mathcal{Y}) - I(\mathcal{Z};\mathcal{X})$ and area **3**=$I(\mathcal{Z};\mathcal{X}) + I(\mathcal{Z};\mathcal{Y}) - I(\mathcal{Z};\mathcal{X},\mathcal{Y})$;

Chapter 6

Multimodal analysis approaches

The combination of imaging modalities requires dedicated analysis methods [Dale and Halgren, 2001, Friston, 2009]. This chapter will give a short overview over existing approaches to data integration of multimodal measurements. Most multimodal analysis approaches are combinations of different unimodal methods (or extensions thereof) as presented in chapter 4. When scanning through the multimodal neuroimaging literature there is a general tendency: The more detailed the data is, the simpler are the methods employed. High resolution techniques such as intracranial microelectrode recordings in combination with intrinsic optical imaging [Grinvald et al., 1994, Devor et al., 2003, Dunn et al., 2005, Berwick et al., 2008] or fMRI [Logothetis et al., 2001] typically employ very simple methods such as event-related averaging, spatial averaging and univariate correlation coefficients. The further away from neural activity the measurements are, the more important are models of the data generating process. For instance when using MEG/EEG measurements, it is crucial to know the conductivity profile of the head (see section 4.3) in order to make inference about the locations of the neural oscillators underlying the measurements. Hence especially in the field of non-invasive multimodal studies using fMRI-EEG there has been an increased interest in model based data analyses [Daunizeau et al., 2007, Riera et al., 2005, Valdes-Sosa et al., 2009]. The following sections will give a broad overview over popular multimodal analysis approaches. In analogy to chapter 4 we divide the methods into *supervised*, *unsupervised* and *model-driven* approaches.

In multimodal imaging a rule of thumb is: The more accurate the measurement, the simpler the analysis method used

6.1 Supervised multimodal methods

Often multimodal methods rely on classical supervised methods (see section 4.1). This means that one modality is used as regressor. Alternatively a stimulus variable can serve as regressor in order to extract relevant features (i.e. those that correlate with the experimentally controlled stimulus) from either modality; for instance ERPs extracted from EEG recordings (using stimulus timing information) can serve as regressor to find fMRI correlates thereof [Eichele et al., 2005]. Analogously fMRI

activity patterns associated with a stimulus are used to bias cortical dipole estimates obtained from EEG measurements [Liu et al., 1998, Dale and Halgren, 2001]. Since in many supervised multimodal analyses one modality is used to guide the analysis of the respective other modality, these methods are often categorized as *asymmetric* data analysis approaches, see e.g. [Laufs et al., 2008, Rosa et al., 2010a]. *Symmetric* data integration approaches try to integrate information from multiple modalities simultaneously. But symmetric analyses based on supervised methods are not applicable to all experimental paradigms. In the following we will give a short overview of asymmetric and symmetric approaches using supervised methods for multimodal data fusion.

Two ways of asymmetric data fusion in supervised learning settings

There are two main approaches to supervised *asymmetric* multimodal fusion.

EEG-constrained fMRI analyses Here, features from EEG data are extracted to *temporally* constrain the fMRI analysis. This is usually implemented by extracting the time course of a certain feature in the EEG data and use that as regressor in the GLM design matrix. The two most often used EEG features are:

Amplitude modulations of event-related potentials (ERPs)
EEG epochs are extracted time locked to the presentation of a sensory stimulus; the amplitude of the ERP can be used as regressor in the GLM design matrix [Eichele et al., 2005];

Amplitude modulations of neural oscillations (Band power)
The strength of neural oscillations in a certain narrow frequency band and a given time window is extracted; the time window length is often matched to that of the fMRI sampling rate; this band power time series can be correlated with the BOLD signal [Logothetis et al., 2001, Moosmann et al., 2003], used as GLM regressor [Laufs et al., 2003, Rosa et al., 2010b] or can be used as target variable in multivariate pattern analysis [Martino et al., 2011].

FMRI-constrained EEG analyses Here, information that is primarily reflected in fMRI activity, is used as a *spatial* constraint for estimating dipole locations based on EEG data. A prominent example is *fMRI-constrained current density imaging* [Liu et al., 1998, Dale et al., 2000] which has already been proposed in [Dale and Sereno, 1993]. Activation patterns obtained from fMRI measurements are assumed to reflect neural activation. Constrained source density imaging uses these patterns as spatial priors to bias the solution of EEG/MEG dipole estimates towards the fMRI activation patterns. While early implementations assumed stationary covariances between cortical sources, this approach has been extended to account for

time-varying spatial constraints [Liu and He, 2008]. In practice fMRI-constrained current density imaging is implemented by estimating an fMRI activation pattern using standard SPM methodology (see section 4.1). This SPM is assumed to reflect the current source covariance modeled in the matrix R in eq. (4.18).

Limitations

Asymmetric supervised methods require unimodal feature selection In the classical supervised learning setting, one of the two modalities is used to extract a univariate regressor. This seemingly negligible limitation makes classical supervised analysis methods difficult to use for multimodal data. Each modality is actually multivariate. Of course classical GLM type analyses can incorporate more than one EEG regressor, see e.g. [Rosa et al., 2010b]; but then the analysis is mass-univariate, meaning that single voxel time series are analyzed separately – and multivariate activation patterns are neglected in the analysis [Lange et al., 1999]. Although the extension of the GLM to multivariate data *and* multivariate regressors is straightforward, this approach has not been investigated in the context of multimodal data integration. Another important aspect is that each modality reflects certain aspects of brain activity that other modalities are blind to (see areas 1 and 2 in figure 5.1). When using asymmetric approaches it is likely that during feature selection of the 'primary' modality (the one that is used to bias the estimate of the other modality) one selects features that are not reflected in the other (secondary) modality at all; trying to find these features in the secondary modality can thus lead astray the scientific interpretation. An example are synchronized oscillations between brain regions. Synchronization measures require a high temporal resolution as they are based on millisecond time differences between oscillations recorded at different positions. These synchronization phenomena are unlikely to be reflected in the slow oscillations of voxel activations in fMRI signals. Unless synchronization between regions imposes additional energy demands on both brain regions, synchronized on the time scale of hemodynamic activation, using synchronization based regressors should not yield any significant activation in the fMRI signal.

In asymmetric multimodal data fusion a regressor is extracted from one modality to bias the activity estimate in another modality

Symmetric supervised methods require a model of neural activity Symmetric integration approaches try to overcome this problem by analyzing modalities simultaneously rather than biasing one towards the other. A symmetric data fusion approach based on supervised learning methods is taken in [Fazli et al., 2011]. The authors use a multivariate pattern classifier on EEG and NIRS data separately. The output of each classifier is then used to train another classifier to predict a stimulus label from the multimodal EEG-NIRS data. This kind of symmetric analysis is not applicable to experimental paradigms studying *intrinsic* neural activity, such as epileptic activity or ongoing oscillations reflecting resting state activity, for a review on paradigm-less multimodal studies see [Salek-Haddadi et al., 2003]. In

Symmetric supervised data integration requires stimulus information

principle supervised methods can be applied here – but this requires manual feature selection e.g. by an experienced neurologist in order to construct a regressor. If there is no model of neural activity and no expert knowledge available for extracting features manually, feature extraction becomes difficult in supervised learning settings. Wrong feature selection can reduce the amount of mutual information between modalities significantly and thus have detrimental effects on subsequent analyses [Ostwald et al., 2010]. If there is no model about the underlying data generation process, which could guide manual or supervised feature extraction, unsupervised methods can be used to automatically extract features.

6.2 Unsupervised multimodal methods

Unsupervised methods extract structure in a data set in an explorative manner. A major problem with unsupervised learning methods for multimodal neuroimaging is the data fusion process. The unsupervised methods in chapter 4.2 are inherently unimodal. Thus when using unsupervised methods on multimodal data often each modality is analyzed separately. When applying PCA or ICA to each modality separately, it is not clear how the components in one modality relate to those of another modality – the first PC in one modality does not necessarily correspond to the first PC in another modality. Data integration is especially difficult if an unsupervised method does not result in an ordered set of components, like ICA. Many approaches extract unimodal features using unsupervised methods from one modality. The extracted features are then used as regressors in a GLM based SPM [Eichele et al., 2005, Debener et al., 2006]. Three examples of solutions to this data integration problem are sketched in the following:

Integration of unimodal analyses [Eichele et al., 2008]:

1. Perform unsupervised analysis separately on each modality.
2. Compute the correlation between single components to find pairs of components across modalities.

Joint analyses of concatenated data [Moosmann et al., 2008]:

1. Reduce the dimensionality of each modality separately by PCA.
2. Concatenate the data of all modalities into one matrix.
3. Apply unsupervised methods to the joint data matrix.

Reciprocally constrain each modality by other modality [Yang et al., 2010]:

1. Perform unimodal temporal ICA separately on EEG data.
2. Extract the band power of one IC to construct a regressor.

3. Use the IC regressor to estimate SPM based fMRI activation pattern.
4. Use the fMRI activation pattern to guide the EEG dipole estimation.

These data integration approaches combine the results from unimodal unsupervised analyses. Other unsupervised data integration approaches are specifically designed for multimodal data and thus do not require a second step of data fusion. Many unsupervised multimodal data integration approaches are special cases of a statistical learning method called *canonical correlation analysis* (CCA) [Hotelling, 1936]. A detailed derivation of CCA is described in chapter 7. Related methods are for instance partial least-squares [Wold, 1985, Geladi and Kowalski, 1986] or multivariate linear regression; also the GLM (see section 4.1) can be cast as a form of CCA [Henson, 2000]. For a comprehensive introduction to CCA and its relationship to other methods see [Borga, 1998]. The advantage of these multimodal unsupervised methods is that they automatically extract the relevant features and integrate the data in one optimization step.

Limitations

Unimodal unsupervised methods are difficult to interpret When using unsupervised methods for data integration, it can be difficult to relate the activation profiles in one modality to those found in another modality. In general when analyzing multimodal data with unimodal methods, the correspondence between the components found in multiple modalities cannot be guaranteed. In extreme cases, ICA would find in one modality independent components that do not show clear correlations with any of the components in another modality. Or a component in one modality is correlated to more than one component in another modality.

Temporal alignment requires a model of neurovascular dynamics Multimodal unsupervised fusion requires each sample in one modality to have a corresponding sample in the other modality – such that single samples can be correlated. In order to get corresponding samples from a high resolution (in time) electrophysiological measurement and a low resolution fMRI time series, one can take similar approaches as in the supervised settings above: Extraction of stimulus locked epochs, as done in [Correa et al., 2010], or extraction of band power time series with temporal bins that have the same duration as the fMRI image acquisition, as done in [Martínez-Montes et al., 2004, Murayama et al., 2010, Bießmann et al., 2011a]. Having the same number of temporally aligned samples is not enough. Unsupervised multimodal analyses require that simultaneous samples in each modality be correlated. For combined electrophysiology and hemodynamic measurements this is not the case. The HRF delays the fMRI signal by several seconds. Temporal alignment is usually done by convolving the electrophysiological time series with the same canonical HRF as in classical supervised analysis methods, see fig. 4.1. This approach does not take into

account the spatiotemporal dynamics and hence can fail to detect certain aspects of neural activity. A solution to this problem will be proposed in chapter 7 and 8.

Only activity reflected in both modalities is found Multimodal unsupervised data fusion is based on the assumption that all modalities reflect a common underlying latent variable. CCA for instance aims to find those representations of each modality that maximizes the correlation between multiple modalities . This implies that those aspects of brain activity that are reflected in one modality only (see figure 5.2, area 1 and 2) will not be captured by CCA. In practice however proper preprocessing helps to avoid these cases: Extracting the fast oscillations of the neurophysiological signal by computing the band power in a preprocessing step captures those aspects that are not directly reflected in the BOLD signal and reduces them to the relevant quantity – changes in membrane potentials on the level of neural ensembles. Similarly proper selection of voxels on the side of fMRI data prevents from biasing the result towards highly active voxels containing e.g. vessels on the cortical surface.

6.3 Model driven data fusion

Detailed biophysical models for multimodal data fusion are proposed e.g. in [Riera et al., 2005, Daunizeau et al., 2007, Plis et al., 2010, Bojak et al., 2011]. In short, the models usually employ forward models for electrical and vascular signals as presented in section 4.3. The modeling typically starts at a hypothetical quantity $x(t)$ representing the combined postsynaptic activity of a neural ensemble in a given voxel. This synaptic activity depends on the synaptic activity at previous time points and potentially also on input from other voxels. Additive noise $\varepsilon_i(t)$ drives the system. Synaptic activity is then translated via appropriate forward models into unimodal measurements. The dynamics of the vascular and electrophysiological forward models are represented in systems of stochastic non-linear differential equations (SDEs). Following the line of thought in [Valdes-Sosa et al., 2009] EEG/fMRI models are particular cases of State Space Models (SSMs) [Kalman, 1960]

$$\dot{x}(t) = f(x(t), s(t), \Phi) + \mu + \Lambda\dot{\varepsilon}_i(t) \tag{6.1}$$

$$y(t) = g(x(t), \Phi, s(t)) + \varepsilon_e(t), \tag{6.2}$$

where ε_e denotes external measurement noise. Equation (6.1) is often called *state equation*, containing a set of SDEs describing the dynamics of the state vector $x(t)$. A state refers to the common underlying quantity driving electrophysiological and hemodynamic activity and is often referred to as *ensemble of postsynaptic potentials* [Valdes-Sosa et al., 2009]. The vector $s(t)$ describes external inputs which influence the system, such as sensory stimulation, ascending neuromodulatory influences or

ongoing oscillations. The set of parameters specifying the forward models is contained in Φ. In case of a combined EEG-fMRI model these parameters include the volume conductivity profile of the head (for the EEG forward model, see section 4.3) and the parameters of the vascular forward model (section 4.3). The vector $\varepsilon_i(t)$ models random inputs following a Gaussian distribution, with mean μ and covariance matrix Λ. Equation (6.2) is called *observation equation* and describes how the (multimodal) measurements are determined by the states $x(t)$. Note that the measurements are taken at discrete time steps t, while the underlying stochastic differential equations are continuous in time. SSMs are well known in control theory since the 1960s and have become popular in the neuroimaging community mainly under the name *Dynamic Causal Modeling* [Friston et al., 2003]. Solving these forward models can be difficult. Analytical solutions to these equations do not exist, they have to be solved numerically. The solution to equations (6.1) and (6.2) can be approximated using *local linearization* [Riera et al., 2007, Valdes-Sosa et al., 2009].

Limitations

Which level of abstraction is the best? A general problem with model based data fusion is that it is difficult to assess how complex a model should be. A good model should be detailed enough to make hypotheses about the neural processes of interest but at the same time the model should be simple enough to be computationally tractable. Obviously it is impossible to model brain activity on a realistically small scale. Neither does one have the measurements to build a sufficiently detailed model, nor is it computationally tractable. Strictly speaking if one wants to describe a neuron exactly, one needs to take into account the state of each ion channel and many metabolic processes inside the cell. With current computer technology and the limited knowledge we have about the neural and vascular anatomy this seems ambitious even for point neuron[1] models, not to speak of more complex neuron models. A typical voxel in a multimodal analysis contains thousands of neurons. Even if it would be technically feasible to place multiple electrodes into every single cell – for a typical voxel size this would amount to a few thousand electrophysiology channels per voxel – there remains the problem of scalability of these models. Efficient solutions for large scale models exist, see e.g. [Valdes-Sosa et al., 2009], but for a data set that would offer a realistic chance to test the validity of the model assumptions, these approximations would be infeasible to compute as well. Abstraction is needed to make the problem tractable. Most studies simplify the problem and do not model single cells but rather net activity of cell ensembles. This approximation seems reasonable. The net dipole strength can be measured with intracranial elec-

[1]Point neurons are modeled as a point in space, without accounting for the cell's morphology; the morphology however does play an important role for the integration of inputs and subthreshold computations; point neuron models are called phenomenological as they reproduce many phenomena of real neurons' activity profiles – without a biologically realistic structure;

trodes. Importantly these measurements can be combined with measurements of fMRI signals [Logothetis et al., 2001]. This allows to test hypotheses involving the model quantities. Other popular simplifying assumptions about the neurovascular coupling mechanisms are:

1. The HRF has an isotropic gaussian profile in space [Sirotin et al., 2009]
2. The spatiotemporal dynamics of the hemodynamic response is independent in space and time [Daunizeau et al., 2007]

Both assumptions speed up analyses considerably. However there is converging empirical evidence speaking against them [Aguirre et al., 1998, Yacoub et al., 2006, Berwick et al., 2008, Bießmann et al., 2011a].

Realistic models are difficult to test empirically Realistic models are only useful as long as they can be tested empirically. If the model quantities cannot be measured nor falsified they are of limited use for scientific research [Popper , 1934]. Some more detailed models for multimodal data fusion explicitly model inhibitory and excitatory neural populations, see e.g. [Bojak et al., 2011]. This is biologically plausible – a single neuron can only have positive spiking rates, never negative activity. For a balanced network it is necessary to have both inhibition and excitation [Brunel, 2000]. However when modeling excitation and inhibition independently, these quantities and the model predictions associated with them have to be testable empirically. Measuring (morphologically or immunohistochemically identified) inhibitory and excitatory cells independently in vivo and ideally in combination with hemodynamic measurements is technically very challenging and to the best of my knowledge there are no labs that have tried to accomplish this task[2].

Summary

Taken together most approaches to multimodal neuroimaging analysis are based on unimodal methods and can be categorized in analogy to chapter 4. The main problems associated with the respective analysis approach are:

- *Supervised methods* find only activity correlated with a univariate regressor
- *Unsupervised methods* can be difficult to interpret in a multimodal context
- *Model driven methods* are difficult to compute/validate for realistic models

[2]Inhibitory interneurons are typically very small and thus intracellular recordings are difficult to obtain. Classifying inhibitory and excitatory neurons only on the basis of their action potential shapes is technically feasible but as there are very fast spiking excitatory neurons it is often difficult to make a clear cut distinction.

Chapter 7

Data-driven multimodal analysis

This chapter will introduce a new data driven approach to multimodal data fusion, *temporal kernel canonical correlation analysis* (tkCCA). In order to meet the requirements for multimodal data acquired with state of the art measurement devices the method was designed to fulfill three criteria:

1. *Multi-modality* – without unimodal feature selection
2. *Scalability* – in particular to fMRI data (>10^4 voxels but few data points)
3. *Model free* – coupling parameters should be estimated from data

This chapter will give a step-by-step introduction to tkCCA. The first step is a recapitulation of the classical statistical learning technique *canonical correlation analysis* (CCA) [Hotelling, 1936]. The use of CCA is motivated by the first requirement, as CCA is specifically designed for multimodal data. In its classical form CCA is computationally costly for high dimensional data such as fMRI signals. Thus in order to meet the second requirement, scalability, we will introduce in a second step the *kernelized* version of CCA, *kernel* CCA (kCCA) [Fyfe and Lai, 2000, Akaho, 2001, Melzer et al., 2001]. Although initially designed for data with non-linear dependencies, another advantage of kCCA is that it is very efficient for high dimensional data, especially if the number of data points is much lower than the number of dimensions. Neurophysiological and hemodynamic activity is not correlated instantaneously but they are related via complex neurovascular coupling mechanisms. The temporal dynamics of this coupling are often modeled by a canonical HRF, see fig. 4.1. This temporal profile of the HRF is hypothesized to be independent of the spatial dynamics of neurovascular coupling. These assumptions are a convenient approximation, however there is empirical evidence against them. The HRF varies between brain regions [Aguirre et al., 1998] and cortical depth [Yacoub et al., 2006]. High resolution optical imaging studies suggest that the local fine structure of cortical vascularization gives rise to spatiotemporal hemodynamics that cannot be captured by a combination of an independent spatial and temporal component [Berwick et al., 2008]. Accounting for this spatiotemporal variability can

increase the prediction accuracy of neurophysiological activity and reveal areas of activation that separable models would fail to detect [Lu et al., 2007]. For a detailed analysis of neurovascular coupling mechanisms it is preferable to refrain from these rather restrictive assumptions. In order to meet the third requirement, model-free analysis, we extended kCCA such that it can account for arbitrary spatiotemporal coupling mechanisms. The resulting algorithm, *temporal kernel canonical correlation analysis* (tkCCA) does not require the usual assumptions about neurovascular spatiotemporal dynamics, instead it estimates the optimal dynamics from the data. For the sake of simplicity all algorithms in this chapter will be illustrated with simple toy data examples. Chapter 8 will then give an overview over the applications of tkCCA to more realistic artificial data and data from real experiments.

7.1 Canonical Correlation Analysis (CCA)

Suppose we have simultaneously recorded measurements of different multivariate modalities $[x_1, x_2, \ldots, x_T] = X \in \mathbb{R}^{U \times T}$ and $[y_1, y_2, \ldots, y_T] = Y \in \mathbb{R}^{V \times T}$. Let us assume that the data are artifact free and both modalities reflect neural activity. How should one integrate these data in a common framework? One modality has U dimensions, the other has V dimensions. It is not clear how each dimension in one modality is related to another dimension in the other modality. The only assumption one can rely on is that both modalities reflect a common underlying process z, in the case of multimodal neuroimaging neural activity. Figure 7.1 depicts the graphical model for this multimodal setting. One could apply PCA on each modality or on a joint data set and continue with analyzing only the strongest principal components of the data. In fact this is often done in multimodal analysis frameworks as a preprocessing step [Moosmann et al., 2008]. However the directions of maximal variance *within* one modality might not be the ones reflecting the signal of interest. Figure 7.2 illustrates this scenario with toy data. In the toy data setting, the common underlying variable, a sine wave, has low variance in each modality. A standard dimensionality reduction in fMRI preprocessing discards the lowest variance directions in the data. Here, joint PCA would discard the very component of interest. In general when applying unimodal unsupervised methods there is no guarantee that the ith component of one modality reflects the same process as the ith component in another modality. This is even worse when using ICA, where the components cannot be not ranked according to the amount of variance explained like in PCA. A simple solution is offered by CCA, which finds new representations for simultaneously measured modalities such that the modalities are maximally correlated. The idea behind CCA is that the representation of the data that maximizes the correlation between the modalities x and y reflects the common underlying process z best.

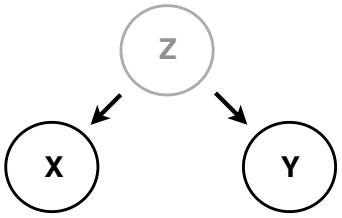

Figure 7.1: Graphical model for CCA; a common hidden variable Z, representing the true neural activity, is reflected in two modalities: X, the measurements of neurophysiological spectrograms and Y, measurements of high dimensional fMRI time series;

Derivation We assume centered data such that $\sum_{i=1}^{T} x_i = 0$ and $\sum_{i=1}^{T} y_i = 0$. CCA learns for each modality a feature weighting w_x, w_y, such that the *canonical correlation*[1] ρ

$$\rho = \frac{w_x^\top C_{xy} w_y}{\sqrt{w_x^\top C_{xx} w_x \; w_y^\top C_{yy} w_y}} \tag{7.1}$$

between the data in X and Y is maximized. Equation 7.1 contains the empirical *auto-covariance matrices* of each data source

$$\frac{1}{T} XX^\top = C_{xx} \in \mathbb{R}^{U \times U} \quad \text{and} \quad \frac{1}{T} YY^\top = C_{yy} \in \mathbb{R}^{V \times V}, \tag{7.2}$$

the empirical estimate of the *cross-covariance matrix*

$$\frac{1}{T} XY^\top = C_{xy} \in \mathbb{R}^{U \times V} \tag{7.3}$$

and the first *canonical directions* $w_x \in \mathbb{R}^U$ and $w_y \in \mathbb{R}^V$ of X and Y, respectively. The numerator contains the *cross-variance* of the data in its new representation, i.e. after projecting the data onto w_x, w_y. In the denominator we have a product of the *auto-covariances* within each data set, projected onto the canonical directions. Without this denominator, ρ would merely be a covariance. With the auto-covariance terms, ρ will be scaled into the interval $[-1 1]$. This correlation coefficient is called *canonical* since ρ is invariant with respect to linear transformations of the data. Many extensions to CCA have been proposed, including non-linear versions of CCA [Leurgans

[1]The canonical correlation is equivalent to the concept of *principle angles*, introduced earlier in [Jordan, 1875]. Principal angles quantify the similarity between two subspaces in an euclidean space.

et al., 1993], CCA using stochastic processes [Fyfe and Leen, 2006] and sparse versions of CCA [Fyfe, 2006, Wiesel et al., 2008, Hardoon and Shawe-Taylor, 2011]. It could also be shown that the CCA framework can be easily extended to more than two multivariate variables [Kettenring, 1971]. CCA has found widespread application in multivariate statistics [Anderson, 1958], signal processing [Larimore, 1990, Bach and Jordan, 2002], data mining [Bie and Cristianini, 2004, Bießmann and Harth, 2010], artificial intelligence [Kim et al., 2007] and neuroscience [Friman et al., 2002, Hardoon et al., 2007, Macke et al., 2008] to name just a few examples.

The complete solution of CCA are sets of canonical projections $W_x = [w_x^1, \ldots, w_x^k]$ and $W_y = [w_y^1, \ldots, w_y^k]$, $k = \min(U, V)$, one set of projections for each variable. Each W_i spans a subspace in the respective data space of the ith modality that maximizes eq. (7.1), the canonical correlation between the variables. We will first derive the solution for the first pair of canonical directions w_x^1, w_y^1 using the Lagrange multiplier technique. This will lead to a CCA solution based on eigenvalue decompositions (see appendix A.1) of covariance matrices. This formulation naturally extends to a CCA solution for all remaining k canonical directions. We want the first pair of canonical directions w_x, w_y to maximize the covariance between X and Y:

$$\underset{w_x, w_y}{\operatorname{argmax}}\, w_x^\top C_{xy} w_y, \tag{7.4}$$

In order to have a *correlation* instead of a *covariance* we need the right normalization for eq. (7.4). This can be done by requiring the covariance *within* each variable X, Y to be 1, i.e. we maximise eq. (7.4) subject to:

$$w_x^\top C_{xx} w_x = 1, \quad w_y^\top C_{yy} w_y = 1. \tag{7.5}$$

Combining the objective function in eq. (7.4) with the constraints in eq (7.5) yields the Lagrangian

$$\mathcal{L} = w_x^\top C_{xy} w_y - \frac{1}{2}\lambda_x \left(w_x^\top C_{xx} w_x - 1 \right) - \frac{1}{2}\lambda_y \left(w_y^\top C_{yy} w_y - 1 \right). \tag{7.6}$$

As we want to maximise eq. (7.6) we calculate the partial derivatives with respect to w_x and set it to zero:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial w_x} = C_{xy} w_y - \lambda_x C_{xx} w_x &= 0 \\ \rightarrow \quad C_{xy} w_y &= \lambda_x C_{xx} w_x \end{aligned} \tag{7.7}$$

Analogously for w_y:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial w_y} = w_x^\top C_{xy} - \lambda_y w_y^\top C_{yy} &= 0 \\ &\rightarrow \quad w_x^\top C_{xy} = \lambda_y C_{yy} w_y \end{aligned} \tag{7.8}$$

By left multiplying eq. (7.7) with $w_x^\top$ we obtain:

$$w_x^\top C_{xy} w_y = \lambda_x w_x^\top C_{xx} w_x \tag{7.9}$$

and by left multiplying eq. (7.8) with $w_y^\top$:

$$\begin{aligned} w_y^\top w_x^\top C_{xy} &= \lambda_y w_y^\top C_{yy} w_y \\ w_y^\top (C_{yx} w_x)^\top &= \lambda_y w_y^\top C_{yy} w_y \\ w_x^\top C_{xy} w_y &= \lambda_y w_y^\top C_{yy} w_y. \end{aligned} \tag{7.10}$$

We see from eq. (7.9), eq. (7.10) and eq. (7.5) that $\lambda_x = \lambda_y = \rho$. Combining eq. (7.7) and eq. (7.8) in one equation results the generalized eigenvalue equation

$$\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} = \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} \lambda. \tag{7.11}$$

Successive generalized eigenvectors W_x, W_y are orthogonal in the metric defined by the within-modality covariance matrices C_{xx}, C_{yy}. We can thus formulate the CCA solution for all canonical projections in complete analogy to eq. (7.11) as

$$\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} W_x \\ W_y \end{bmatrix} = \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix} \begin{bmatrix} W_x \\ W_y \end{bmatrix} \Lambda. \tag{7.12}$$

where $W_x \in \mathbb{R}^{U \times k}$, $W_y \in \mathbb{R}^{V \times k}$ are the matrices with the respective canonical projections in their columns and

$$\Lambda = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_k \end{bmatrix} \tag{7.13}$$

is the diagonal matrix of all canonical correlations. We can now project the data into the subspace with the highest canonical correlation by multiplying the data with the j columns of W that correspond to the j highest eigenvalues,

$$X' = \sum_{i=1}^{j} e_i w_x^{i\top} X = [w_x^1 w_x^2 \dots w_x^j]^\top X = W_x^\top X \tag{7.14}$$

$$Y' = \sum_{i=1}^{j} e_i w_y^{i\top} Y = [w_y^1 w_y^2 \dots w_y^j]^\top Y = W_y^\top Y. \tag{7.15}$$

Generative Model CCA can be interpreted in a probabilistic framework [Bach and Jordan, 2006]. The generative model underlying the canonical directions is based on a multivariate gaussian latent variable $z \in \mathbb{R}^D$ drawn from a multivariate normal distribution with diagonal covariance matrix Λ

$$z \sim \mathcal{N}(0, \Lambda), \qquad \Lambda \in \mathbb{R}^{D \times D}. \tag{7.16}$$

This is the underlying variable in fig. 8.2 which is reflected in the modalities $x \in \mathbb{R}^U$ and $y \in \mathbb{R}^V$. The *conditional* probability distributions of x given z, and y given z, respectively are given by

$$\begin{aligned} x|z &\sim \mathcal{N}(W_x z, Q_x), W_x \in \mathbb{R}^{U \times D} \\ y|z &\sim \mathcal{N}(W_y z, Q_y), W_y \in \mathbb{R}^{V \times D}. \end{aligned} \tag{7.17}$$

The maximum likelihood solution to eq. (7.17) can be estimated using expectation maximization [Bach and Jordan, 2006]. Extensions to mixtures of non-gaussian distributions such as Dirichlet distributions have been proposed in [Fyfe and Leen, 2006]. Although the generative model perspective is helpful in many cases, the solution to eq. (7.17) involves expectation-maximization across the parameters and is thus computationally costly for multimodal data. Often an analysis is not interested in estimating the full probability density function of the variables. If the goal of an analysis is merely to decode neural activity it is simpler to find for each modality the features w_x, w_y that represent that underlying neural process best. For this task the algebraic perspective in eq. (7.11) is sufficient if a fast and efficient implementation is needed.

Extensions to more than two variables CCA can be extended to more than two variables, see [Kettenring, 1971]. In general if there are N multivariate variables and corresponding centered data matrices $\{X_1, X_2, \ldots, X_N\}$, the basis vectors of the canonical subspace of each variable $\{W_1, W_2, \ldots, W_N\}$ can be found by solving the generalized eigenvalue problem

$$\begin{bmatrix} 0 & C_{12} & \ldots & C_{1N} \\ C_{21} & 0 & \ldots & C_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ C_{N1} & C_{N2} & \ldots & 0 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_N \end{bmatrix} = \begin{bmatrix} C_{11} & 0 & \ldots & 0 \\ 0 & C_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & C_{NN} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_N \end{bmatrix} \Lambda. \tag{7.18}$$

where C_{ij} denotes the covariance matrix between the ith and jth variable.

CCA and mutual information There is a straightforward relationship between canonical correlation and mutual information. If the data is completely characterized by its first two moments, mean and covariance, the mutual information $I(X, Y)$

(see eq. (5.2)) is given by

$$I(X, Y) = -\frac{1}{2} \sum_i \log \left(\frac{1}{(1 - \lambda_i^2)} \right). \tag{7.19}$$

So the mutual information between two modalities X, Y can be computed by summing up the diagonal entries of the new cross-covariance matrix Λ after projecting the data on their canonical directions. Proofs can be found in [Borga, 1998] or [Cover and Thomas, 2006]. This means that, leaving higher order moments aside and assuming elliptical data distributions of X and Y, CCA optimizes the mutual information between two modalities (corresponding to area 3 and 4 in fig. 5.2). At first glance one might think that optimizing mutual information would miss exclusively unimodal information (areas 1 and 2), but include irrelevant features belonging to multimodal artifacts of non-neuronal origin (area 4). In practice however, standard preprocessing techniques (proper artifact correction and unimodal preprocessings such as band power extraction for electrophysiological data) help to avoid these cases.

Why CCA is better than PCA for multimodal data PCA is often used in multimodal analyses for unimodal feature extraction. It is easy to see in a simple toy data example why this can lead to wrong conclusions. Figure 7.2 illustrates the PCA and CCA solution on a two dimensional toy data set, i.e. $x \in \mathbb{R}^2, y \in \mathbb{R}^2$, generated according to:

$$X = \begin{bmatrix} \sqrt{\gamma}\, s(t) \\ \sqrt{1-\gamma}\, \varepsilon_x(t) \end{bmatrix} \qquad Y = \begin{bmatrix} \sqrt{1-\gamma}\, \varepsilon_y(t) \\ \sqrt{\gamma}\, s(t) \end{bmatrix} \tag{7.20}$$

where the underlying common signal $s(t)$ is a sine wave (of arbitrary frequency) and the measurement noise $\varepsilon_x(t), \varepsilon_y(t)$ for each variable is drawn each from a normal distribution with mean zero and standard deviation one. The variance of the signal is controlled by the parameter γ, in our simulations we set $\gamma = 0.1$ in order to have a *low signal variance* in one dimension of X, Y and a *high noise variance* $1 - \gamma$. As shown in fig. 7.2A/B the generative model in eq. (7.20) assigns the low variance signal to the first dimension of X but to the second dimension of Y, the respective other dimension carries the noise. Computing PCA separately on this multimodal data will assign high weights to the dimension of X, Y that exhibit most variance and low weights to those that carry low variance. If the common underlying signal s has a lower variance than the modality specific noise $\varepsilon_x, \varepsilon_y$, PCA will consider the noise to be the most important aspect of the data and the first principal component will reflect not the common signal but only noise. CCA in contrast will maximize not the variance *within* each variable, but rather the correlation *between* X and Y. If the noise within each data source is stronger than the common signal in multiple

modalities CCA clearly yields more sensible results. Figure 7.2C/D show the results of PCA and CCA on the toy data generated according to eq. (7.20); while the first principal component captures only the modality specific noise, the first canonical component reflects only the common signal s.

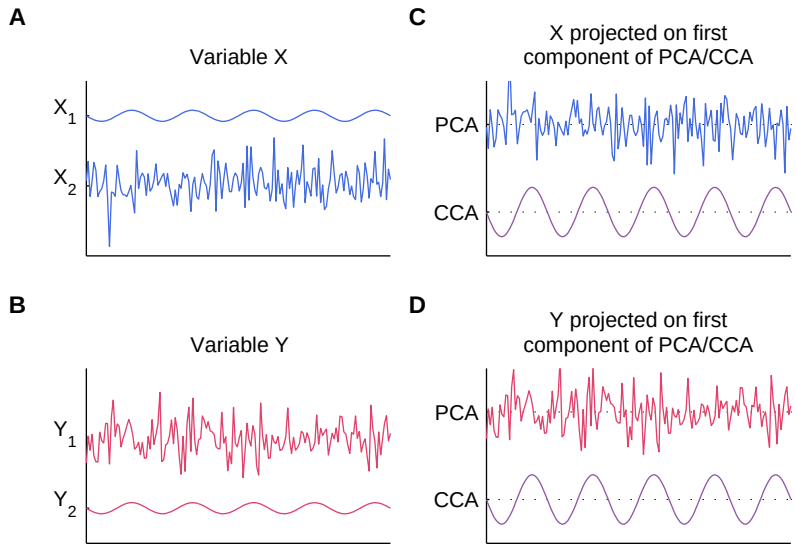

Figure 7.2: Toy data example comparing PCA and CCA for multimodal data analysis; data was generated according to eq. (7.20); **A**: first dimension of X contains high variance noise ε_x and second dimension the low variance signal s; **B**: Same for Y, dimensions for s and ε_y are flipped compared to X; **C**: First principal and canonical component of X; PCA captures only the noise, CCA captures the signal; **D**: Same as C for variable Y;

7.2 Kernel CCA (kCCA)

Young man, in mathematics
you don't understand things.
You just get used to them.
—*John von Neumann*

There are situations in which the (canonical) correlation coefficient fails to detect dependencies between two modalities. A simple example of non-linear dependencies between two univariate variables is a phase shift as illustrated in fig. 7.3. In the example we consider univariate variables only for the sake of simplicity. The variables are generated by

$$x = sin(t) + 0.1\varepsilon, \qquad y = cos(t) + 0.1\varepsilon \tag{7.21}$$

such that the signals are exactly the same but phase shifted by $\pi/2$, as shown in fig. 7.3A. When plotting x against y the two signals will show zero correlation (fig. 7.3B). This example was chosen for illustrative purposes only – of course there are many other methods that could be used to extract the non-linear dependencies in this particular toy data example; e.g. simply correcting for the time lag of $\pi/2$ would remove the non-linearity in this example. In general if there are dependencies between two variables that cannot be captured by linear relationships one can use the *kernel trick* [Aizerman et al., 1964, Schölkopf et al., 1998, Müller et al., 2001] to extract non-linear dependencies from the data. The idea behind the *kernel trick* is to map all data points $x \in \mathbb{R}^U, y \in \mathbb{R}^V$ into a much higher (possibly infinite-dimensional) space $\mathcal{S}$

$$\varphi_x : \mathbb{R}^U \to \mathcal{S}, \qquad \varphi_y : \mathbb{R}^V \to \mathcal{S} \tag{7.22}$$

and solve the solution to a learning problem in that space. The trick is that one never has to compute these mappings explicitly if the objective function can be expressed in terms of inner products in $\mathcal{S}$, denoted $\langle ., . \rangle_{\mathcal{S}}$. All computations can then be done on the kernels K_X, K_Y, where the ith row and jth column are the inner product between the ith and jth data point in the kernel feature space

$$K_{X,ij} = k_x(x_i, x_j) = \langle \varphi_x(x_i), \varphi_x(x_j) \rangle_{\mathcal{S}}, \tag{7.23}$$

where the *kernel function* $k_x(.,.)$ computes the inner product in $\mathcal{S}$. This of course requires $\mathcal{S}$ to have an inner product and a norm. Popular kernel functions that map to a space $\mathcal{S}$ which meets these criteria include

$$\text{Linear kernels} \qquad k(x_i, x_j) = x_i^\top x_j, \tag{7.24}$$

$$\text{Polynomial kernels} \qquad k(x_i, x_j) = (x_i, x_j)^d, \tag{7.25}$$

$$\text{Gaussian kernels} \qquad k(x_i, x_j) = e^{(x_i - x_j)^2/(-2\sigma_x^2)}.$$

In kCCA the original CCA objective in eq. (7.4) becomes

$$\varphi_x^*,\ \varphi_y^* = \underset{\varphi_x,\varphi_y}{\operatorname{argmax}} \operatorname{Corr}(\varphi_x(X), \varphi_y(Y)), \tag{7.26}$$

where

$$\begin{aligned} \varphi_x(X) &= [\varphi_x(x_1), \varphi_x(x_2), \ldots, \varphi_x(x_T)]^\top \\ \varphi_y(Y) &= [\varphi_y(y_1), \varphi_y(y_2), \ldots, \varphi_y(y_T)]^\top \end{aligned} \tag{7.27}$$

are the T data points mapped into $\mathcal{S}$. The intuition behind kernel CCA is that the solutions φ_x^*, φ_y^* have to lie in the space spanned by the data. This means that regarding the data points mapped into $\mathcal{S}$ as basis vectors of a subspace in $\mathcal{S}$, the solution can be expressed as

$$\begin{aligned} \varphi_x^* &= \sum_i^T \varphi_x(x_i)\alpha_i = \varphi_x(X)\alpha, \\ \varphi_y^* &= \sum_i^T \varphi_y(y_i)\beta_i = \varphi_y(Y)\beta, \end{aligned} \tag{7.28}$$

where $\alpha \in \mathbb{R}^T$ and $\beta \in \mathbb{R}^T$ and the index $i \in \{1, \ldots, T\}$ runs over all training data points. The entries $\alpha_i,\ \beta_i$ contain the contribution of each data point i to the solution $\varphi_x^*,\ \varphi_y^*$. Solving kernel CCA thus means finding the $\alpha,\ \beta$ vectors that maximize eq. (7.26). This is all that is needed in order to express the solutions $\varphi_x^*,\ \varphi_y^*$ in $\mathcal{S}$. Following eq. (7.28) new samples $x_i,\ y_i$ can be projected onto their canonical directions in kernel space by

$$\begin{aligned} \langle \varphi_x(x_i), \varphi_x^* \rangle_{\mathcal{S}} &= \langle \varphi_x(x_i), \sum_j^T \varphi_x(x_j)\alpha_j \rangle_{\mathcal{S}} = \sum_j k_x(x_i, x_j)\alpha_j = K_x\alpha \\ \langle \varphi_y(y_i), \varphi_y^* \rangle_{\mathcal{S}} &= \langle \varphi_y(y_i), \sum_j^T \varphi_y(y_j)\beta_j \rangle_{\mathcal{S}} = \sum_j k_y(y_i, y_j)\beta_j = K_y\beta, \end{aligned} \tag{7.29}$$

where $\alpha_j,\ \beta_j$ are the jth entry of $\alpha,\ \beta$, respectively. It turns out that the solution to kernel CCA, the coefficients α and β, can be obtained just as easily as the solution to standard CCA in the previous section. Looking at eq. (7.29) and the original CCA objective in eq. (7.4) we see that $w_x^\top C_{xy} w_y = 1/T\ \ w_x^\top X Y^\top w_y$ is equivalent to

$$\langle \varphi_x(X), \varphi_x^* \rangle_{\mathcal{S}}^\top \langle \varphi_y(Y), \varphi_y^* \rangle_{\mathcal{S}} = \alpha^\top K_x K_y \beta. \tag{7.30}$$

The Lagrangian in eq. (7.6) thus becomes

$$\mathcal{L} = \alpha^\top K_x K_y \beta - \frac{1}{2}\lambda_x \left(\alpha^\top K_x^2 \alpha - 1\right) - \frac{1}{2}\lambda_y \left(\beta^\top K_y^2 \beta - 1\right). \tag{7.31}$$

So the solution of kernel CCA is exactly the same equation except that the canonical projections $w_x,\ w_y$ are replaced by $\alpha,\ \beta$ and the covariance matrices $C_{xy},\ C_{xx},\ C_{yy}$ are replaced by the respective kernel products $K_x K_y,\ K_x^2,\ K_y^2$. In complete analogy to the CCA case we can find the solution to eq. (7.31) by taking the partial derivatives and setting them to zero. The resulting equations can be rearranged and the coefficients $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_T]^\top$ and $\beta = [\beta_1, \beta_2, \ldots, \beta_T]^\top$ can then be obtained analogously to eq. (7.12) as solution to the generalized eigenvalue problem

$$\begin{bmatrix} 0 & K_X K_Y \\ K_Y K_X & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \rho \begin{bmatrix} K_X^2 & 0 \\ 0 & K_Y^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \tag{7.32}$$

Statistical consistency of kernel CCA has been shown in [Fukumizu et al., 2007]. An important advantage of kCCA is the ability to detect non-linear dependencies. Figure 7.3 shows that while the linear correlation coefficient is zero, after mapping the data through a gaussian kernel function the data are perfectly correlated.

Non-linear kernel functions offers a powerful means to estimate non-linear dependencies [Gretton et al., 2005]. But the visualization of the canonical directions in feature space is often difficult. Simple toy data examples in two dimensions are easy to visualize, see fig. 7.3. For most real data sets in which kernel methods are being used this is more difficult. Especially in the context of high dimensional data such as fMRI signals often linear kernels are sufficient. Non-linear kernels do not necessarily lead to an improvement as to the accuracy at which a data set can be predicted from a stimulus [Bergstrand et al., 2009]. And more importantly linear kernels offer the advantage that the solutions to eq. (7.26) $\varphi_x^*,\ \varphi_y^*$ can be easily visualized in the input space of the respective modality as a linear expansion of data points:

$$\varphi_x^* = X\alpha, \quad \varphi_y^* = Y\beta. \tag{7.33}$$

where the coefficients α can be obtained by solving eq. (7.32). Note that while the solution to linear kCCA is the same as linear CCA, the computation of this solution is much more efficient in kernel space as the kernel of the data is only of size $L \times L$, while the covariance matrices would be of size $U \times U$ and $V \times V$. In the case of fMRI data this offers a significant computational speed-up.

Regularization Real data has a limited number of samples. Within a given data set there will be spurious correlations that do not reflect the true coupling dynamics but rather noise fluctuations. When fitting any model given a limited number of data points, it is thus important to control the complexity of the solution. Otherwise the

results will be *overfitted* to the training data set – and they will not generalize to new data. Controlling the complexity of the solution is called *regularization* and can be implemented by requiring the solutions to the generalized eigenvalue problem α, β to have a small euclidean norm. This additional constraint can be incorporated in the standard kCCA constraint by simply adding a ridge of height κ_x, κ_y onto the respective kernel matrix

$$\alpha^\top K_x^2 \alpha + \kappa_x I = 1 \tag{7.34}$$

$$\beta^\top K_y^2 \beta + \kappa_y I = 1, \tag{7.35}$$

where I denotes the identity matrix of the same size as $K_x,\ K_y$. Plugging these constraints in equation (7.31) yields a new Lagrangian

$$\mathcal{L} = \alpha^\top K_x K_y \beta - \frac{1}{2}\lambda_x \left(\alpha^\top K_x^2 \alpha + \kappa_x I - 1\right) - \frac{1}{2}\lambda_y \left(\beta^\top K_y^2 \beta + \kappa_y I - 1\right). \tag{7.36}$$

Computing the derivative, setting it to zero and rearranging the equations, the solution for regularized *kernel* CCA can be formulated as:

$$\begin{bmatrix} 0 & K_X K_Y \\ K_Y K_X & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \rho \begin{bmatrix} K_X^2 + \kappa_x I & 0 \\ 0 & K_Y^2 + \kappa_y I \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{7.37}$$

So regularizing the kCCA solution can also be achieved by simply adding a ridge on the denominator of the generalized eigenvalue equation (7.32). The choice of the right hyper-parameter is crucial. Typically one optimizes the regularizers $\kappa_x,\ \kappa_y$ by using cross-validation [Lemm et al., 2011] or permutation based approaches as in [Bießmann et al., 2010b].

The pre-image problem Equation (7.29) shows that new data points can be easily evaluated in kernel space. So one can compute canonical correlations in $\mathcal{S}$. But often one is also interested in looking at φ_x^* and φ_y^*, to see which features kernel CCA picked to maximize the canonical correlation. Unfortunately the space in which these solutions are living is potentially infinite dimensional and thus difficult to interpret. Sometimes it can be helpful to find the mapping $\varphi_x^{-1},\ \varphi_y^{-1}$ that brings φ_x^*, φ_y^* from $\mathcal{S}$ back into the respective input space $\mathbb{R}^U,\ \mathbb{R}^V$

$$\varphi_x^{-1} : \mathcal{S} \to \mathbb{R}^U, \quad \varphi_y^{-1} : \mathcal{S} \to \mathbb{R}^V \tag{7.38}$$

Finding the right $\varphi_x^{-1},\ \varphi_y^{-1}$ is called the *pre-image problem*. The pre-image of linear kernels is given by eq. (7.33). Solving the pre-image problem for the non-linear case is not as simple. Various methods for visualizing the solution of non-linear methods in input space have been proposed [Schölkopf et al., 1999, Mika et al., 1999, Kwok and Tsang, 2003, Bakir et al., 2004] but these methods do not always yield satisfactory results.

While an exhaustive treatment of kernel methods is beyond the scope of this thesis, there are two features of kernel methods that are useful for the practitioner to keep in mind:

1. Non-linear problems can become linear in kernel feature space
2. A kernel algorithm operates only on similarities between data points

The first point highlights the power of kernel methods but also emphasizes the need for a reasonable choice of a kernel; non-linear dependencies in the original data space often become linear in kernel feature space – but not always the solution in kernel space is easy to visualize and interpret in the original data space. The second aspect has its positive and negative sides, too: Working on inner products in feature space means that the computational cost of a kernel based algorithm will scale with the number of data points – and not with the dimensions of the feature space nor with the dimensions of the input space. In case the dimensionality of a data set is much larger than the number of data points, using kernel methods can considerably speed up an algorithm. But if the number of data points is larger than the number of dimensions, kernel methods can be much slower than equivalent non-kernelized algorithms.

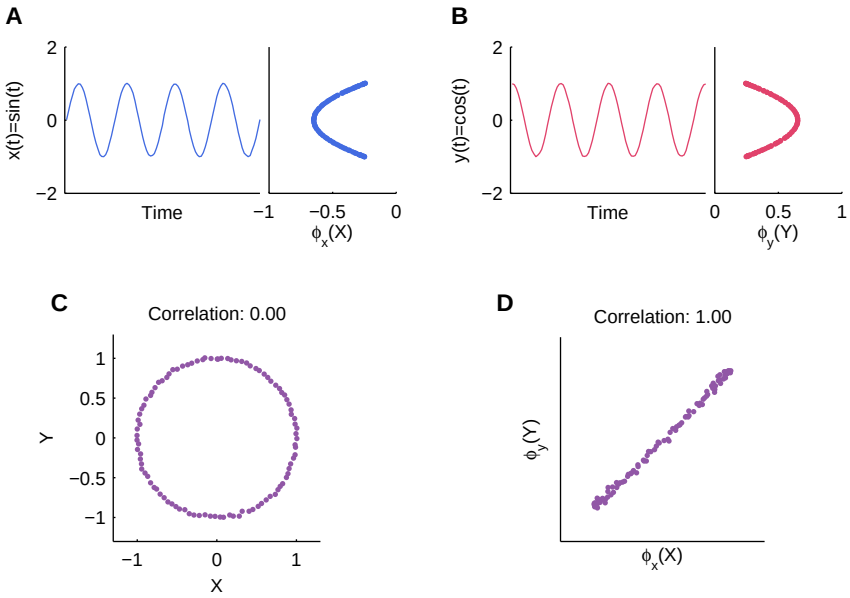

Figure 7.3: Illustration of kernel CCA on toy data generated according to eq. (7.21); kernel function was a gaussian kernel with $\sigma = 1$, see also eq. (7.24); **A**: Time series of first modality $x(t) = sin(t)$; for each value of X the corresponding value $\varphi_x(X)$ in the kernel feature space is plotted to the right of the time series plot; **B**: Same as A for $Y = cos(t)$; **C**: The standard linear canonical correlation coefficient between X and Y is zero; **D**: First canonical component of X and Y obtained by kCCA plotted against each other; in kernel space, the two signals are perfectly correlated;

7.3 Temporal kCCA (tkCCA)

It's just a temporal embedding?
I can't believe nobody tried this before.

—*Nicole Krämer*

In CCA the canonical correlation, see eq. (7.1), between two signals is maximized. When the two signals are not *instantaneously* correlated but with some time lag τ then classical CCA will fail to find sensible canonical directions. The toy data example in figure 7.3 shows such a scenario. Non-linear kernel CCA could find the dependency between the variables nonetheless. But often there is no need to use non-linear kernel functions. In the toy data example the data is actually linearly dependent: If one knew the correct time lag of $\pi/2$ one could shift the cosine signal back in time and the signals would be perfectly correlated. In general when multiple modalities are not coupled instantaneously the dependencies can be described as a convolution. Neurovascular coupling mechanisms for instance can not be modeled by a mere phase shift between neural and hemodynamic signal, as in the case of the toy data example of kCCA. The hemodynamic response to a neural impulse in a given neural population is a spatiotemporal pattern in voxel space. If one knows this pattern, one can use it to deconvolve the voxel time series and obtain an estimate of the underlying neural signal. *Temporal kernel canonical correlation analysis* (tkCCA) [Bießmann et al., 2010b] computes this convolution from high dimensional multimodal data streams and thereby compensates for non-instantaneous coupling between modalities.

Derivation The idea behind tkCCA is simple: As the correlations between X and Y are not instantaneous, the canonical correlation coefficient will be small for simultaneous samples. By shifting one data source back and forth in time (relative to the other data source) by some relative time lag $\tau = [-T, -T+1, \ldots, 0, \ldots, T-1, T]$, the correlation will become larger for the correct time lag. Ideally we would like to have a time lag dependent weighting $w_x(\tau) \in \mathbb{R}^U$ for one data source (the one shifted in time) that can compensate for delays or convolutions between data sources. The canonical convolution $w_x(\tau)$ should be such that there are high weights for those time lags and features that maximize the correlation with the time series in Y. If $w_x(\tau)$ is estimated correctly the convolution of X with $w_x(\tau)$ will compensate for data-source specific differences in temporal dynamics. In the case of the neurovascular coupling example the canonical convolution $w_x(\tau)$ will lowpass filter the neural activity and bring the signal in temporal alignment with the hemodynamic activity [Bießmann et al., 2010b, Murayama et al., 2010]. In tkCCA the time resolved

canonical convolution $w_x(\tau)$ and the canonical direction w_y are optimized such that:

$$\underset{w_x(\tau),w_y}{\operatorname{argmax}} \left(\frac{\sum_\tau (w_x(\tau)^\top X_\tau)^\top Y w_y}{\sqrt{\sum_\tau (w_x(\tau)^\top X_\tau X_\tau^\top w_x(\tau)) w_y^\top Y Y^\top w_y}} \right) \tag{7.39}$$

where X_τ denotes the data in X shifted by a time lag τ and $w_x(\tau)$ is the canonical convolution of X at time lag τ. For $\tau = 0$ equation (7.46) reduces to the standard CCA, as in eq. (7.4). In general tkCCA not only optimizes the canonical correlation for simultaneous samples, but over a short time window specified by the range of τ. To maximize eq. (7.39), we embed the data set X into a higher dimensional space where the additional dimensions are occupied by copies of the original data, each shifted in time by a different τ:

$$\tilde{X} = \begin{bmatrix} X_{\tau=-T} \\ \vdots \\ X_{\tau=T} \end{bmatrix} \in \mathbb{R}^{M(2T+1)\times L}. \tag{7.40}$$

By using the same time-delay embedding for the $w_x(\tau)$,

$$\tilde{w}_x = \begin{bmatrix} w_x(\tau = -T) \\ \vdots \\ w_x(\tau = +T) \end{bmatrix} \in \mathbb{R}^{M(2T+1)\times 1} \tag{7.41}$$

we reduce the optimization problem in equation (7.39) to an ordinary CCA problem in the embedded space that can be solved as in eq. (7.12) by maximizing the CCA objective using the embedded data in $\tilde{X}$:

$$\underset{\tilde{w}_x,w_y}{\operatorname{argmax}} \frac{\tilde{w}_x^\top \tilde{X} Y^\top w_y}{\sqrt{\tilde{w}_x^\top \tilde{X}\tilde{X}^\top \tilde{w}_x w_y^\top Y Y^\top w_y}}. \tag{7.42}$$

The tkCCA solution is thus the CCA solution in the embedded space. Using linear kernel CCA we can express the τ-dependent filter $w_x(\tau)$ as an expansion of the data points in $\tilde{X}$ and the respective other canonical projection w_y as an expansion of the data points in Y in complete analogy to equation (7.33)

$$\begin{aligned} \tilde{w}_x &= \begin{bmatrix} w_x(\tau = -T) \\ \vdots \\ w_x(\tau = T) \end{bmatrix} = \tilde{X}\alpha, \\ w_y &= Y\beta. \end{aligned} \tag{7.43}$$

Or, remembering the structure of $\tilde{X}$ in eq. (7.40), the delay-dependent canonical convolution $w_x(\tau)$, can be recovered equivalently as expansion of the α vector using the respective parts of $\tilde{X}$:

$$w_x(\tau) = X_\tau \alpha. \tag{7.44}$$

Solving tkCCA thus requires to find the optimal dual coefficients α and β. Including a regularizing ridge $\kappa_x I$ and $\kappa_y I$ for each modality, the tkCCA solution α and β can be obtained by solving the generalized eigenvalue problem

$$\begin{bmatrix} 0 & K_{\tilde{X}} K_Y \\ K_Y K_{\tilde{X}} & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \rho \begin{bmatrix} K_{\tilde{X}}^2 + \kappa_x I & 0 \\ 0 & K_Y^2 + \kappa_y I \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \tag{7.45}$$

where $K_{\tilde{X}} = \tilde{X}^\top \tilde{X}$ is the linear kernel of the temporally embedded data matrix. For non-linear kernel functions of course $K_{\tilde{X}}$ and K_Y are computed according to the kernel function chosen. Note that the kernel computed from the embedded data $\tilde{X}$ is of the same size as the kernel K_X. Thus after the kernel $\tilde{K}_X$ has been computed tkCCA is computationally not more demanding than normal kCCA. Another advantage of tkCCA is that cross-validation procedures or bootstrap re-samplings can operate on the already computed kernel, which can speed up the re-samplings considerably. Since tkCCA just leads to a single eigenvalue equation of the type given by eq. (7.37), neither the canonical correlation ρ nor the corresponding vectors α and β depend on τ. Nonetheless tkCCA can reveal the temporal dynamics between the data in their respective canonical subspace. In analogy to a standard cross-correlogram for univariate time series, we can compute a time lag dependent canonical correlogram $\rho(\tau)$ by

$$\rho(\tau) = \frac{w_x(\tau)^\top X_\tau, w_y^\top Y}{\sqrt{w_x(\tau)^\top X_\tau X_\tau^\top w_x(\tau) w_y^\top Y Y^\top w_y}} \tag{7.46}$$

$$= \frac{\alpha^\top K_\tau K_Y \beta}{\sqrt{\alpha^\top K_\tau^2 \alpha \cdot \beta^\top K_Y^2 \beta}}, \tag{7.47}$$

where $K_\tau = k_x(X_\tau, X_\tau)$ is the kernel computed from X_τ, so in the case of linear kernels $K_\tau = X_\tau^\top X_\tau$. The canonical correlogram $\rho(\tau)$ reflects the temporal dynamics in kernel feature space $\mathcal{S}$ between the two variables and can be considered as a multivariate generalization of the univariate cross-correlogram. An example is shown in figure 7.4D. It is worth mentioning that the canonical correlogram is different from a usual correlogram between two one-dimensional time signals in that it is always positive [2]. This is because CCA is ignorant with respect to the sign of the correlation; possible negative correlations between the signals are encoded in the relative sign of the canonical directions. So, to interpret the results of tkCCA, one always has to take into account both the canonical correlogram $\rho(\tau)$ and the corresponding canonical directions $w_x(\tau)$ and w_y. The temporal resolution of the correlogram is determined by the spacing of the τ-values, in particular it can be higher than the temporal resolution of the signals in Y. Another important aspect is that while the shape of the

[2] Since the $\rho(\tau)$ are not directly calculated by kCCA, but by eq. (7.46), this is not strictly true, i.e. only up to noise fluctuations.

canonical directions in kernel feature space might be difficult to visualize, the temporal dynamics as reflected in $\rho(\tau)$ can easily be computed in kernel feature space using eq. 7.47. In case the data is high dimensional this can offer a considerable speedup of computations, if only the temporal dynamics of correlations (in kernel feature space) are of interest. As tkCCA can be cast as a standard kCCA problem, the statistical consistency shown in [Fukumizu et al., 2007] also applies to tkCCA.

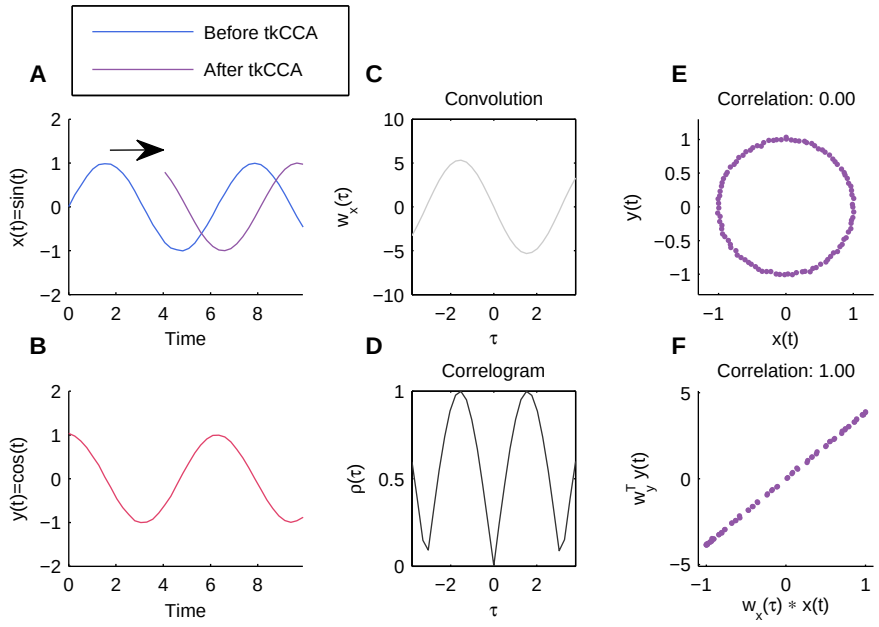

Figure 7.4: Toy data example for tkCCA on same data as in previous chapter; **A:** First modality $x(t) = sin(t)$ (in blue) and first canonical component obtained by tkCCA (in purple); **B:** Second modality $y(t) = cos(t)$; **C:** Canonical convolution $w_x(\tau)$; **D:** canonical correlogram $\rho(\tau)$; **E:** Raw signals $x(t), y(t)$ are not correlated; **F:** After convolving the signal $x(t)$ with $w_x(\tau)$, the resulting signal $x(t) * w_x(\tau)$ is perfectly (anti-)correlated with $y(t)$; the convolution with $w_x(\tau)$ aligned the two signals (see red time series in *B* and purple time series in *A*), but as CCA is inherently blind to the sign of the correlation, the canonical correlation is negative;

7.4 Overview and comparison

This chapter introduced a data driven approach to multimodal analysis starting from CCA over kernel CCA to temporal kernel CCA. In order to differentiate these methods it is helpful to remember to which kind of data they are applicable. Table 7.1 lists

Dimensionality	Dependencies	Appropriate Method		
		CCA	kCCA	tkCCA
low	instantaneous linear	√	√	√
high	instantaneous linear	×	√	√
low/high	instantaneous non-linear	×	√	√
low/high	non-instantaneous non-linear	×	×	√

Table 7.1: Comparison of data-driven multimodal methods;

the different kinds of dependencies exhibited between modalities and the appropriate CCA method.

If the data is low dimensional and has instantaneous linear dependencies, standard CCA will find the correct solution. The computational cost of standard CCA is quadratic in the number of dimensions, while kernel CCA has quadratic cost with respect to the number of data points. Most computers nowadays can deal with covariance matrices up to a few thousand dimensions – fMRI data can easily have a few ten thousand dimensions but the number of samples rarely exceeds a few hundred. Here, linear kernel CCA offers a more robust and much faster way of computing canonical projections compared to standard covariance matrix based methods. Next to the computational speed-up kernel CCA can detect non-linear dependencies – whereas standard CCA is restricted to linear dependencies. If the data is *high dimensional* or exhibits non-linear dependencies, kernel CCA is better than standard CCA.

Most multimodal neuroimaging data has non-instantaneous couplings – simultaneously recorded samples are not correlated. Importantly the coupling relationship is not simply a phase shift, as in our toy data example. In this case, kernel CCA will fail to find the optimally correlated (between modalities) representation for each modality. Temporal kernel CCA will find the dependency structure for non-instantaneously coupled data. In principle, standard CCA could also be used for estimating spatiotemporal neurovascular coupling dynamics. But if one wants to estimate a multivariate *convolution* for fMRI data, the estimation problem increases by a factor of T, the number of time lags; for a ten thousand dimensional fMRI data set with a temporal resolution of 1Hz a convolution of length 20 seconds thus requires to solve a 200,000 dimensional problem. In this scenario, when we need to account for convolutions between high dimensional data streams, tkCCA has a clear advantage. Moreover using appropriate kernels, also non-linear dependencies can be captured by tkCCA. This enables tkCCA to extract causal relationships between multivariate and non-linearly coupled time series [Wu et al., 2011].

Chapter 8

Model-free analysis of neurovascular coupling

In the previous chapter a purely data driven approach to multimodal neuroimaging analysis, termed temporal kernel canonical correlation analysis (tkCCA), was developed. The advantage of tkCCA is that it does not require the usual restrictive assumptions about neurovascular coupling mechanisms. Most classical analysis approaches to fMRI data assume that every voxel reflects neural activity with the same temporal dynamics – typically modeled as a canonical HRF, see fig. 4.1. In contrast tkCCA can account for non-separable coupling dynamics, that is multivariate patterns in voxel space than change across time lags. This chapter will show applications of tkCCA to model-free estimation of neurovascular coupling parameters. A special focus will be placed on the spatiotemporal dynamics of the hemodynamic response, in particular spatiotemporally non-separable coupling dynamics. Using data from simultaneous recordings of intracranial electrophysiological and hemodynamic activity we applied tkCCA to estimate high dimensional spatiotemporal deconvolutions that predict neurophysiological spectrograms from voxel volume time series around the recording electrode. The scientific question that can now be addressed is whether spatiotemporally non-separable dynamics of the BOLD response contain neural information. If this is the case then tkCCA will predict neural activity better than standard methods, which are based on the spatiotemporal separability assumption. This chapter is structured as follows: First section 8.1 introduces the data model used for the simulations. Then section 8.2 presents the analysis settings in which tkCCA was applied. Section 8.3 describes how the different classical analysis models have been computed for comparison with tkCCA. Thereafter section 8.4 shows applications of tkCCA on artificial data and comparisons with standard methods. Section 8.5 summarizes the artifact cleaning procedure for simultaneous electrophysiology and fMRI data that was implemented to run online during an experimental session. Finally section 8.6 will present applications of tkCCA to real data and comparisons with standard methods.

Functional Form	Matrix Form	Meaning
$z(t)$	$z \in \mathbb{R}^{1\times T}$	Effective neural activity
$x(f,t)$	$X \in \mathbb{R}^{F\times T}$	Neural band power (F frequency bands, T time points)
$y(s,t)$	$Y \in \mathbb{R}^{S\times T}$	Vectorized fMRI signal (S voxels, T time points)
ε		Gaussian i.i.d. noise $\mathcal{N}(0,1)$

Table 8.1: Nomenclature: Neural spectrograms $x(f,t) = X \in \mathbb{R}^{F\times T}$ are measured at time $t \in \{t_1, t_2, \ldots, t_T\}$ and in frequency bin f. Hemodynamic signals at each time point t and at each location in voxel space $s \in \{s_1, s_2, \ldots, s_S\}$ is given by the voxel activation $y(s,t) = Y \in \mathbb{R}^{S\times T}$.

8.1 A simple generative model for neurovascular coupling

A simple generative model for multimodal neuroimaging data measured at a certain location s_0 in the brain is depicted in figure 8.1 and can be described as follows: Brain activity $z(t)$ underlying neurophysiological measurements as well as the BOLD signal is reflected instantaneously in neural spectrograms with a frequency band specific weighting $\alpha(f)$ plus some measurement noise

$$x(f,t) = \alpha(f)z(t) + \varepsilon_x. \tag{8.1}$$

Since not all aspects of neural signals are reflected in the BOLD signal we will refer to $z(t)$ as *effective neural activity*, meaning that part of neural activity that is reflected in both hemodynamic and neural signals. The intuition behind the frequency dependency $\alpha(f)$ is that some frequency bands are more correlated to the BOLD signal than others [Logothetis et al., 2001]. The spatiotemporal dynamics of the hemodynamic response can be modeled as a linear filter $H(s,\tau)$:

$$\hat{y}(s,t) = \sum_{\tau,s} H(s,\tau)\, z(t+\tau) + \varepsilon_y. \tag{8.2}$$

The parameter $\tau \in \{0, 1, \ldots, \tau_{\max}\}$ denotes the time lag of the hemodynamic response relative to a neural event at time point t. When using this model in the following we will make, without loss of generality, three simplifying assumptions:

Band power representation of neural activity Neural activity $x(f,t)$ is not modeled in its raw form as voltage fluctuations. Instead we consider only the slower dynamics of amplitude fluctuations of oscillations within a certain frequency range. The summed variance within a frequency band reflects the activity and thus energy demands within a neural population. The exact shape of membrane potential fluctuations can be important e.g. for spike sorting when looking for single cell activity.

But band power modulations are closer to the neural mass-activity on the scale of a voxel and preserve most of the information in the neural signal that is relevant for modeling the BOLD signal.

Neurovascular coupling modeled as linear filter The motivation for this model assumption was twofold: Using linear filters one can deal with high-dimensional data efficiently while being able to visualize the filters easily. The framework presented is however not restricted to the linear case. In general one might use arbitrary non-linear kernel functions for tkCCA, the interpretations of the results would however be more difficult. A more physiologically motivated reason to not use non-linear kernel functions is that the temporal dynamics of the hemodynamic response are well captured by a linear filter [Martindale et al., 2003].

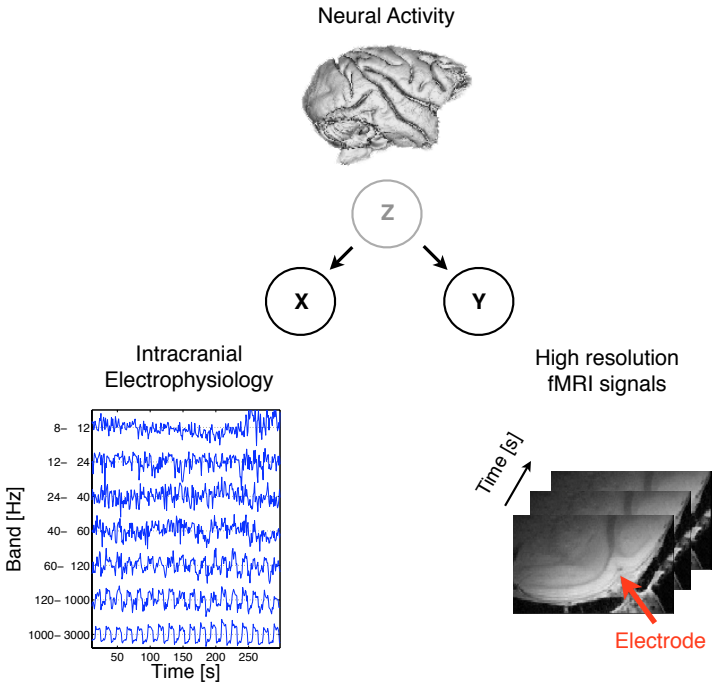

Figure 8.1: Illustration of the data analyzed in this chapter, for experimental details see appendix B or [Murayama et al., 2010]; neural activity Z is instantaneously reflected in electrophysiological measurements X (*left*); a convenient representation of electrophysiological signals are amplitude fluctuations of oscillations in different frequency bands; note the band specific sensitivity to visual stimulation, high frequencies of neural activation are strongly modulated by the visual stimulus; time series of fMRI images Y (*right*) were recorded in the entire primary and parts of secondary visual cortices;

Same temporal resolution across modalities In the examples presented in the following the modalities are assumed to have the same temporal resolution. In general the temporal resolution of tkCCA depends only on the temporal resolution of the faster sampled signal. One can use a sliding window approach as in [Bießmann et al., 2010b] to compute canonical convolutions with a temporal resolution higher than that of the fMRI image acquisition.

8.2 Predicting neural activity from fMRI and vice versa

TkCCA computes for each modality a new representation such that the two modalities are maximally correlated. As the modalities of interest, electrophysiological oscillations and fMRI signals, are coupled via complex neurovascular coupling mechanisms, the temporal coupling dynamics have to be accounted for by equipping one canonical direction with a temporal dimension. When applying tkCCA to simultaneous recordings of neural and hemodynamic activity one has to decide which modality should be temporally embedded, see eq. (7.40); for this modality tkCCA yields a time resolved canonical direction. As we can only use *relative time lags* between sources only one canonical direction at a time is allowed to have a temporal resolution. Both approaches temporally align the two signals, but the choice of the modality to be embedded determines what conclusions can be drawn from the tkCCA result. For the two modalities in our data set there are two options. If one embeds the electrophysiological band power signal $x(f,t)$, tkCCA computes a canonical convolution $w_x(\tau) \in \mathbb{R}^{F\times\tau_{\max}}$ and a canonical projection $w_y \in \mathbb{R}^S$. This allows for investigating the temporal coupling dynamics of different frequency bands of neural signals with static voxel patterns. The other option is to embed the fMRI signal $y(s,t)$ such that tkCCA computes a spatiotemporal deconvolution $w_y(\tau) \in \mathbb{R}^{S\times\tau_{\max}}$ and a frequency band projection $w_x \in \mathbb{R}^F$. This allows for analyzing how the spatiotemporal coupling dynamics of the hemodynamic response in voxel space is related to a static neural spectrogram pattern. Both analysis settings are illustrated in figure 8.2:

A *Predicting voxel patterns from time windows of neural spectrograms*

In the first setting (see fig. 8.2A) an estimate $z_e(t)$ of effective neural activity is obtained by convolving the electrophysiological spectrograms $x(f,t-\tau)$ with the canonical convolution $w_x(\tau)$. Along with the convolution $w_x(\tau)$ tkCCA estimates a canonical direction $w_y \in \mathbb{R}^F$ in voxel space. Projecting the fMRI signal onto its canonical direction w_y yields an estimate of effective neural activity $z_f(t)$ based on the fMRI time series $y(s,t)$

$$\begin{aligned} z_e(t) &= \sum_\tau w_x(\tau)^\top x(f,t-\tau) \\ z_f(t) &= w_y^\top y(s,t). \end{aligned} \tag{8.3}$$

The correlation between $z_e(t)$ and $z_f(t)$ is the canonical correlation. The time window used for the convolution is indicated by a grey frame in fig. 8.2A, the time point which is predicted is indicated as a grey vertical line at $t = 0$. Here the temporal dynamics of neurovascular coupling modeled as $H(s,t)$ in eq. (8.2) are reflected in the temporal structure of $w_x(\tau)$ while the spatial dynamics are expressed in the canonical projection w_y. The spatial activation pattern w_y in voxel space is invariant over time; an advantage of this setting is that one can determine the temporal dynamics of neurovascular coupling with respect to different frequencies of neural oscillations. The convolution $w_x(\tau)$ effectively aligns the neural spectrograms with the fMRI time course, similar to a convolution of a stimulus with a standard canonical HRF (see also fig. 8.2). Examples of $w_x(\tau)$ estimated on real data are depicted in figure 8.10.

B *Predicting neural signals from time windows of fMRI time series*

Hemodynamic activity at time points $t + \tau$ can be used to predict neural activity at time point t (see fig. 8.2B):

$$\begin{aligned} z_e(t) &= w_x^\top x(f,t) \\ z_f(t) &= \sum_\tau w_y(\tau)^\top y(s, t+\tau). \end{aligned} \tag{8.4}$$

Here the hemodynamic response pattern $w_y(\tau)$ reflects the inverse of the coupling dynamics $H(s,t)$ modeled in the data generating model in eq. (8.2); w_x reflects the frequency dependency pattern $\alpha(f)$. As this setting gives insights into the spatiotemporal dynamics of the hemodynamic response a special focus will be placed on this approach in the following application examples.

The choice of the time lags τ used for the embedding is important: Neural signals are assumed to be the source of fMRI signals. Thus it would make little sense to use electrophysiological signals at $t+\tau$ for predicting fMRI signals at time t unless there is some periodicity in both signals, e.g. due to a stimulus. Similarly it would not make sense to predict neural activity at time t from fMRI signals at time points $t-\tau$. The physiologically most sensible choice of time lags is depicted in figure 8.2. When predicting fMRI signals at time point t it is reasonable[1] to take into account neural signals at $t - \tau$ where $\tau < 20s$. Analogously when predicting neural signals at time point t from fMRI data it is reasonable to include data at time lags $t + \tau, \tau < 20s$.

[1]The hemodynamic response function decays within approximately 20s [Martindale et al., 2003]

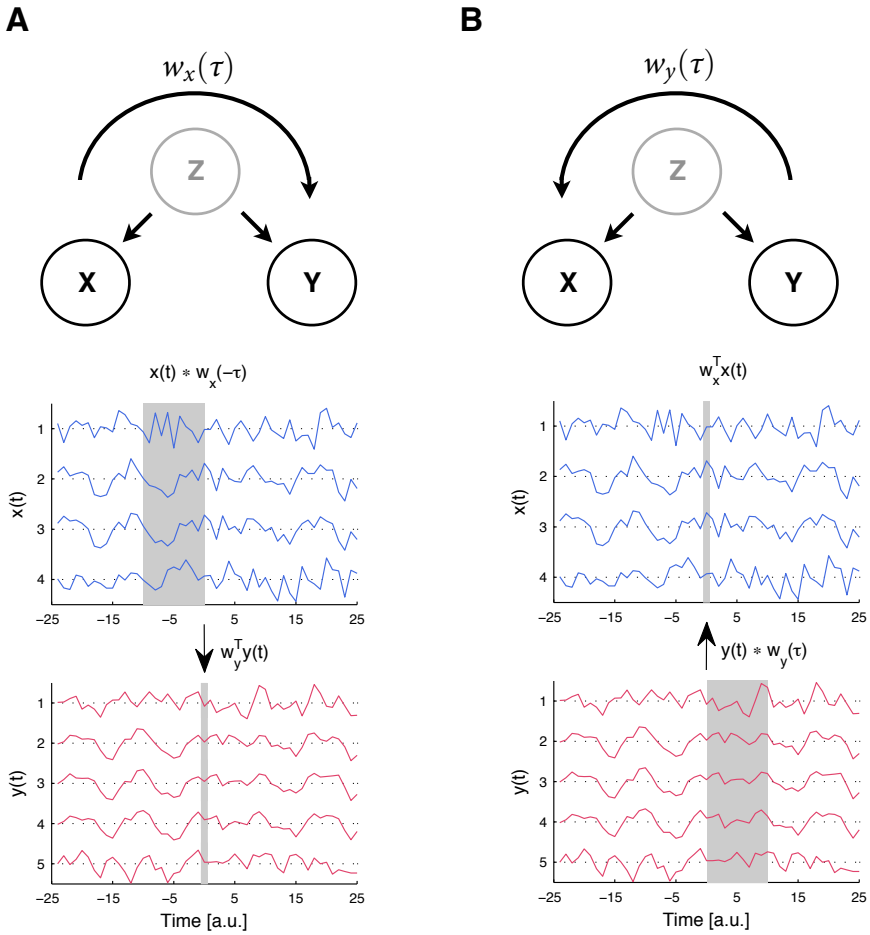

Figure 8.2: Illustration of two approaches to model-free neurovascular coupling analysis using tkCCA; **A**: Time-frequency convolutions $w_x(\tau)$ of neural signals at time points $t-\tau$ (grey area) can predict a voxel pattern w_y at time point t (grey vertical line); **B**: Spatiotemporal deconvolutions $w_y(\tau)$ of fMRI signals at time points $t+\tau$ (grey area) predict spectrograms of neural activity at time point t (grey vertical line);

8.3 Comparisons with other methods

We compared tkCCA with standard methods introduced in chapter 4 in simulations and real data experiments. In the comparisons we consider here only the setting in fig. 8.2B and predict neural activity from fMRI signals. For each method m we compare the tkCCA deconvolution $w_y^*(\tau)$ with a spatiotemporal deconvolution $w_y^m(\tau)$ constructed according to the assumptions underlying that analysis method, see fig. 8.3. Almost[2] all standard methods assume spatiotemporal separability of the hemodynamic response. According to this assumption the spatiotemporal deconvolutions $w_y^m(\tau) \in \mathbb{R}^{S\times N_\tau}$ can be written as the product of two independent functions $w_y^m \in \mathbb{R}^{S\times 1}$ and $w_\tau \in \mathbb{R}^{1\times N_\tau}$ of space and time, respectively,

$$w_y^m(\tau) = w_y^m w_\tau. \tag{8.5}$$

The voxel pattern w_y^m describes the spatial dynamics of the hemodynamic response $H(s,\tau)$ and the temporal dynamics are captured by w_τ. In agreement with the assumption of a canonical HRF the temporal neurovascular coupling dynamics w_τ were the same in all models. All separable models were constructed as follows: An optimal temporal filter w_τ and multivariate voxel pattern w_y^{mv} were obtained from the non-separable estimate $w_y^*(\tau)$ using singular value decomposition (SVD), see e.g. [Strang, 2009]. Examples of w_τ obtained from real data are plotted in fig. 8.12A. The temporal filter w_τ was used for all separable models, the spatial filters were estimated for each model separately. For doing so, we first create a univariate regressor $z_e(t)$ of neural activity from the multivariate spectrograms as described in eq. (8.4). The frequency combination w_x used was the optimal one computed by tkCCA. This univariate regressor was then convolved with the optimal temporal filter w_τ to account for the temporal coupling dynamics. The different spatial dynamics w_y^m were estimated as follows:

1. *Mass-univariate unimodal methods* (see section 4.1, eq. (4.3))
   The mass-univariate spatial filter w_y^{mu} weights each voxel at position s by a factor w_{ys}^{mu} according to its correlation with the neural regressor $z_e(t)$:

   $$w_{ys}^{mu} = \mathrm{Corr}(z_e(t), y(s,t)) \tag{8.6}$$

2. *Multivariate unimodal methods* (see section 4.1, eq. (4.6))
   The multivariate spatial filter w_y^{mv} is the linear combination of voxels that maximizes the correlation between fMRI signals and neural regressor $z_e(t)$:

   $$w_y^{mv} = \underset{w_y}{\mathrm{argmax}}\, \mathrm{Corr}\left(z_e(t), w_y^\top y(s,t)\right) \tag{8.7}$$

   where w_y^{mv} is the spatial factor obtained by a SVD of $w_y^*(\tau)$.

[2] Exceptions are e.g. [Aguirre et al., 1998, Lu et al., 2007, Mourão-Miranda et al., 2007]

3. *Unsupervised unimodal methods* (see section 4.2)
The PCA spatial filter is estimated according to the PCA objective in eq. (4.9):

$$w_y^{pca} = \underset{w_y}{\operatorname{argmax}} \operatorname{Var}\left(w_y^\top y(s,t)\right) \tag{8.8}$$

Note that for the estimation of w_y^{pca} the electrophysiological activity is not taken into account.

4. *Spatial average model*
As a control condition we computed the spatially averaged fMRI signal in the spherical ROI mask around the recording electrode:

$$w_y^{iso}(s,\tau) = \frac{1}{S} w_\tau \tag{8.9}$$

where S denotes the number of voxels within the spherical ROI. The deconvolution $w_y^{iso}(\tau)$ is equivalent to the spatial smoothing often performed as a preprocessing step in fMRI analyses.

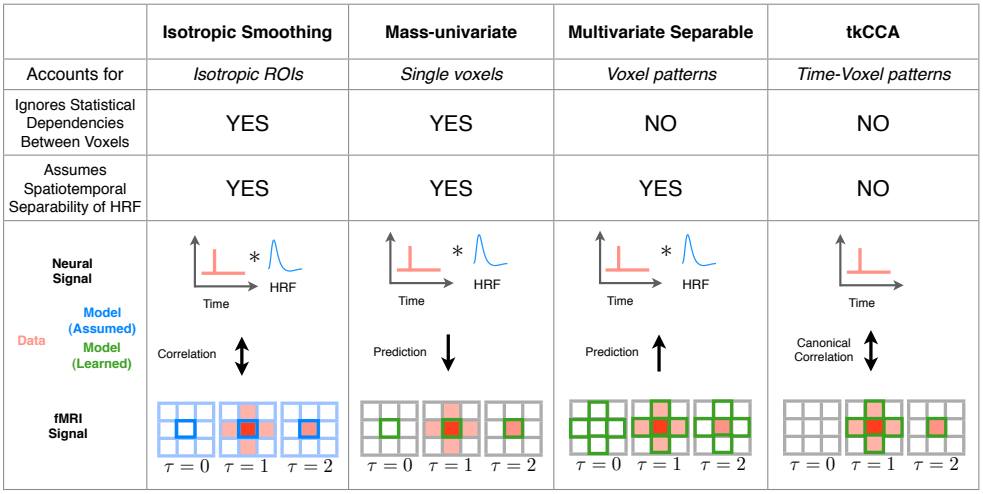

Figure 8.3: Comparison of classical methods and tkCCA; in red a neural impulse and corresponding fMRI signal in 9 voxels at τ=0,1,2 samples after the neural event; in space-time separable models a neural regressor is convolved with a canonical HRF and combined with a static spatial activation pattern – the spatial patterns are invariant over time and the HRF identical for all voxels; *spatial smoothing* weights voxels equally; *mass-univariate* approaches assume statistical independence of voxels and analyze single voxel time series separately; *multivariate separable* models predict neural signals from multivariate voxel patterns; in contrast to space-time separable models tkCCA estimates a non-separable spatiotemporal hemodynamic activation pattern.

8.4 Applications of tkCCA to simulated data

The following simulations illustrate tkCCA and how it compares to standard methods on artificial data. In later chapters tkCCA is used to estimate the spatiotemporal dynamics of hemodynamic activation, according to the setting illustrated in fig. 8.2B. Also the simulations will be based on this setting. The simulated data contains two different time series designed to match the main characteristics of real multimodal neuroimaging data: Variable $x(f,t)$ represents the neural band power in different frequency bands f and time bins t at an electrode position s_0. Variable $y(s,t)$ denotes the fMRI signal at location s in a two dimensional image patch. For the sake of simplicity we assume only univariate hidden variables representing a neural process $z(t)$ which is reflected in both modalities. Following the graphical model depicted in fig. 8.2 the data is generated from this *effective* neural activity $z(t)$, here simply gaussian noise[3], drawn from $\mathcal{N}(0,1)$. From this hidden effective neural activity $z(t)$ electrophysiological measurements $x(f,t)$ are generated by

$$x(f,t) = \big(1-\gamma\big)\alpha(f)z(t) + \gamma\,\varepsilon_x \tag{8.10}$$

and hemodynamic measurements $y(s,t)$ are generated by

$$y(s,t) = \big(1-\gamma\big)\sum_{\tau} H(s,\tau)z(t-\tau) + \gamma\,\varepsilon_y. \tag{8.11}$$

Note that the noise $\varepsilon_x, \varepsilon_y$ is independently drawn but the noise amplitude was controlled by the same parameter γ. The hemodynamic response function $H(s,\tau)$ was constructed such that its spatiotemporal dynamics were non-separable in space and time. This is physiologically plausible – given the complexity of cortical microvasculature (see fig. 3.2) and a variety of factors influencing the BOLD response it is unlikely that neurovascular coupling mechanisms are spatiotemporally separable. Experimental evidence for non-separable hemodynamic response patterns can be found in e.g. [Berwick et al., 2008]. A simple non-separable hemodynamic response was constructed as the sum of two separable functions, one with a fast transient positive response in an $+$-shaped voxel pattern and another one reflecting a delayed undershoot in a $\times$-shaped spatial pattern

$$H(s,\tau) = h_+(s)\,h_I(\tau) + h_\times(s)\,h_U(\tau) \tag{8.12}$$

resulting in the dynamics presented in figure 8.4.

[3]We also tested the effect of periodic signals in $z(t)$, such as stimulus induced variations during block stimulation, a common experimental paradigm in fMRI studies; simulations show that tkCCA will find the true coupling dynamics, even in the presence of periodic 'experimental-stimulus-like' oscillations [Bießmann et al., 2010b].

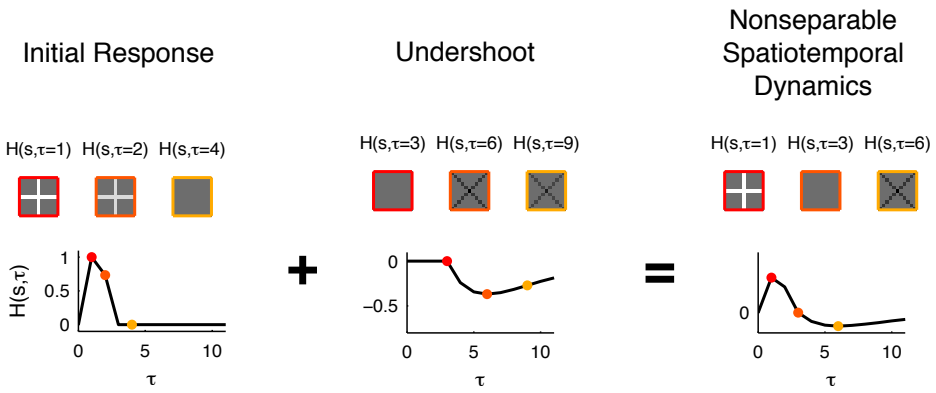

Figure 8.4: Artificial spatiotemporal hemodynamic response function $H(s,\tau)$ used for generating synthetic fMRI time series data according to eq. (8.11); two separable space-time dynamics, an initial response $H_+(s,\tau)$ with a $+$-shaped spatial profile and a later undershoot $H_\times(s,\tau)$ with $\times$-shaped spatial profile were added together yielding a spatiotemporally non-separable hemodynamic response function $H(s,\tau)$;

Parameter	Value
Noise level γ	[0.01 0.2 0.5 0.7]
Number of data points N	[200 400 600]
Number of voxels S	$[11^2\ 31^2\ 51^2\ 101^2]$
length of convolution $\tau_{\max}$	11

Table 8.2: Simulation parameters

Simulation parameters In the simulations all relevant parameters were varied in order to test the performance of tkCCA. The parameter values are listed in table 8.2. We explored a parameter range that is realistic in terms of number of data points N, in fMRI experiments typically around 500, and dimensionality of $y(s,t)$, typically the number of voxels is between 1000 and 10000. The length of most canonical HRF models used in standard fMRI analysis is below 20s which corresponds at a temporal resolution of fMRI signals of 0.5Hz to a maximal time lag of $\tau_{\max}$=10 samples.

tkCCA runtime scales quadratically with the number of data points ...

Effect of voxel number and sample size on tkCCA runtime Figure 8.5 shows the runtime of tkCCA for different numbers of data points and dimensions on a standard laptop. As expected from eq. 7.45 the runtime scales quadratically with the number of data points. If needed, this quadratic increase can be compensated for by not using the full kernel matrices but low rank approximations thereof as suggested in [Bach and Jordan, 2002]. Since in most fMRI experiments the number of data points is around 500 (dashed red line in fig. 8.5A) this quadratic scaling is negligible in practice. Note that the voxel number does not influence the actual tkCCA opti-

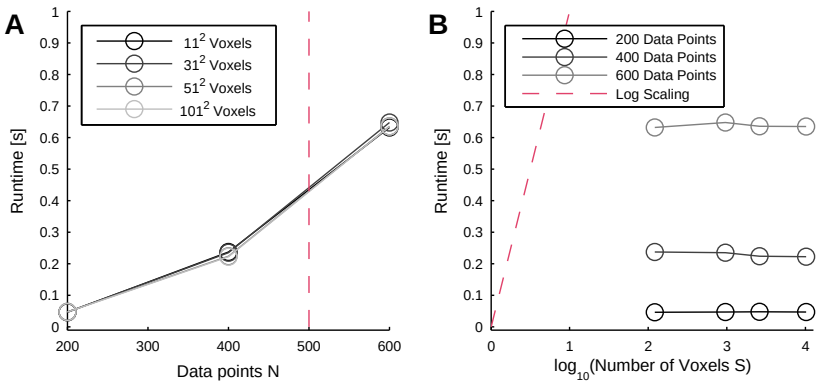

Figure 8.5: Runtime of tkCCA; **A:** Runtime as function of number of data points N; different shades of grey indicate the number of voxels (11^2 31^2 51^2 101^2); all lines fall on each other as the number of voxels does not influence runtime; tkCCA scales quadratically with the number of data points; since fMRI data sets have less around 500 images (dashed red line), this quadratic scaling is negligible in practice; **B:** Runtime of the tkCCA optimization does not increase with dimensionality of the data; different shades of grey indicate the number of data points (200-600); logarithmic scaling with voxel number is indicated by dashed red line;

mization runtime; all lines fall on top of each other in fig. 8.5A. This follows directly from the tkCCA formulation in eq. 7.45. After computing the temporal embedding and the kernel matrices, tkCCA complexity does not depend on the dimensionality of the data anymore. Figure 8.5B shows the same data as in panel A, but here the x-axis shows the number of voxels on a logarithmic scale, from 11^2 to 101^2 voxels. The number of data points is indicated as different shades of grey. Increasing the number of voxels does not influence runtime. Logarithmic scaling is indicated as red dashed line. Note that the true dimensionality of the problem is *number of voxels* $\times$ *number of time lags*; when dealing with 10,000 voxels and 11 time lags this amounts to 10,000$\times$11=1100000. Most fMRI data sets typically have a temporal resolution of 0.25-1Hz, meaning one image acquisition (often called *time of repetition* or TR) takes $4 - 1s$. At a temporal resolution of 0.5Hz 600 samples amount to 20 minutes of fMRI signal. Most standard fMRI data sets contain under 1000 images but more than 10,000 voxels. These data sets can be processed with tkCCA, including cross-validation across several hyper-parameters, in under a minute. This highlights the potential of tkCCA for online applications.

... but tkCCA runtime is independent of the data dimensionality

Accuracy of estimated coupling features Figure 8.6 shows the estimated spatiotemporal deconvolutions $w_y^m(\tau)$ for $\tau = 2$ and $\tau = 11$, each model m in a separate line. The top line depicts the true non-separable spatiotemporal hemodynamic response function $H(s, \tau)$. Results for two different noise levels, $\gamma = 0.2$ and $\gamma = 0.7$,

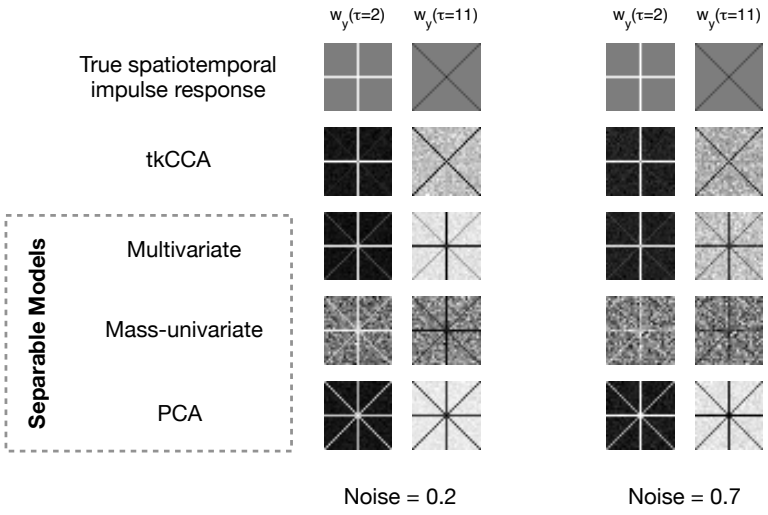

Figure 8.6: Examples of spatiotemporal deconvolutions $w_y(\tau)$ estimated on artificial data generated according to eq. (8.10) and (8.11); for all analysis approaches we plotted the spatial pattern at time lags $\tau = 2$ and $\tau = 11$ for two different noise levels (0.2 and 0.7); the top row shows the true pattern of the coupling dynamics $H(s, \tau)$; number of voxels was 31^2, number of data points used was 200; none of the separable models finds the true spatiotemporal dynamics, while tkCCA recovers the true dynamics even at high noise levels, high dimensional data and few data points;

In contrast to standard methods tkCCA captures non-separable spatiotemporal dependencies

are shown. The spatiotemporal pattern estimated by tkCCA captures the dynamics of the true HRF $H(s, \tau)$. Separable models cannot capture the spatiotemporal dynamics of the hemodynamic response function used for the data generation (see fig. 8.4). The spatiotemporal deconvolutions estimated with methods implicitly assuming spatiotemporal separability are merely a stationary projection w_y in voxel space. In the best case (i.e. low noise conditions, shown in left panels), the deconvolutions of the separable models w_y^{mv}, w_y^{mu} and w_y^{iso} capture a superposition of both dynamics, the transient $+$-shaped pattern and the delayed $\times$-pattern. In contrast tkCCA learns the correct dynamics $H(s, \tau)$. Figure 8.7 shows in the bottom row the similarity between the true coupling $H(s, \tau)$ and the estimated deconvolution averaged across 20 simulations for each parameter setting. Separable models consistently fail to capture the coupling dynamics $H(s, \tau)$. The least flexible model w_y^{iso}, corresponding to spatial averaging of the fMRI signal in combination with a canonical HRF, reflects the coupling dynamics worst. The non-separable deconvolution estimated by tkCCA in contrast learns the correct spatiotemporal dynamics from few data points and high dimensional, noisy data.

Accuracy of neural activity predictions The top row of figure 8.7 shows the canonical correlation between true and estimated neural activity, averaged across 20 simulations. In each panel the top row shows the correlation between neural and estimated neural activity – deconvolved from artificial fMRI data $y(s,t)$ using the model specific spatiotemporal filter $w_y^m(s,\tau)$. TkCCA (red line) predicts neural activity best in all parameter settings. Mass-univariate methods (MU) and spatial averaging (AVG) perform equally well, PCA based deconvolutions predict the underlying neural activity worst.

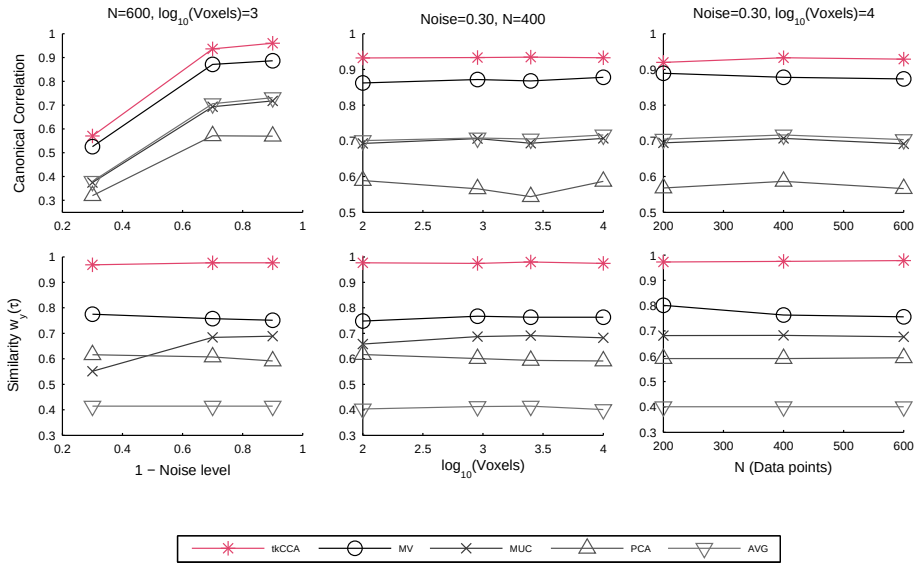

Figure 8.7: Overview of simulation results for (from left to right) different levels of noise, increasing voxel number and increasing number of data points N; top row shows canonical correlation between artificial fMRI time series and neural signals; plotted are results obtained with tkCCA ($*$ symbol), multivariate deconvolutions $w_y^{mv}(\tau)$ ($\circ$ symbol), mass-univariate $w_y^{mu}(\tau)$ ($\times$ symbol), PCA $w_y^{pca}(\tau)$ ($\triangle$ symbol) and the spatial average fMRI signal $w_y^{iso}(\tau)$ ($\triangledown$ symbol); *bottom row* shows the correlation coefficient between the true coupling dynamics $H(s,\tau)$ and the estimated spatiotemporal deconvolution of each model $w_y^m(\tau)$; note that MV, MU, PCA and AVG analyses were provided with optimal temporal dynamics prior to the respective estimation of the spatial dynamics; tkCCA clearly outperforms separable methods in terms of prediction accuracy of neural signals as well as accuracy of the estimated deconvolutions;

8.5 Removal of fMRI scanner gradient artifacts

An important step in multimodal data analysis is artifact removal. When recording electrophysiological signals during fMRI signal acquisition, the magnetic fields of the scanner induce currents in the electrodes which are much stronger than the currents induced by the electromagnetic fields of the brain. Examples of scanner gradient induced artifacts in neurophysiological recordings are shown in fig. 8.8A, C and E. As part of this dissertation a realtime system was developed that removes these artifacts online during an experiment. For artifact removal, the electrophysiological time series is split into epochs whose length matches the length of the acquisition period of one fMRI image (see fig. 8.8 C and E). As the trigger signal sent by the fMRI scanner is not accurate enough we chose to split the epochs according to an empirical threshold on the first derivative of the electrophysiological signal. The steep rise at the beginning of a scanner readout, after the slice selector signal, served as a reference point for temporal alignment of all epochs. The epochs x_i, $i = [1, \ldots, L]$, L being the total number of fMRI volumes acquired, are then stacked as row vectors in a data matrix $X = [x_1^\top, x_2^\top, \ldots, x_L^\top]^\top$. On the resulting data matrix we compute linear kernel PCA. The first principal component is the strongest artifact (see

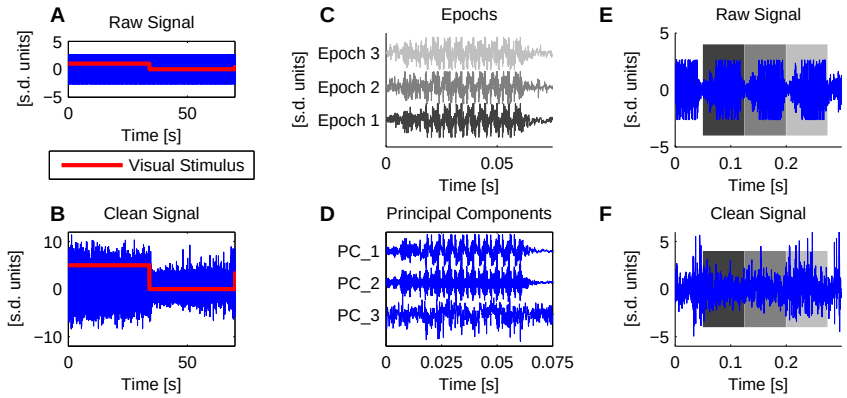

Figure 8.8: Example of PCA based artifact removal from neurophysiological recordings recorded during fMRI data acquisition; **A**: example of raw contaminated time series (in blue) in units of standard deviation; neural responses to visual stimulation (in red) are much smaller than the artifacts and cannot be detected **B**: Clean signal after artifact removal, neural activity robustly follows visual stimulation; **C:** Signal from A (magnified in E) split into single epochs (indicated as background in E/F); note the high amplitude artifacts; **D:** first three principal components computed from epochs in C; **E**: epochs from C concatenated before cleaning; **F**: same as E after subtraction of the PCs plotted in D;

fig. 8.8D). We then retain all but the seven[4] strongest (artifactual) components using eq. (4.12). The cleaning procedure relies on the assumption that the artifacts are temporally aligned in each epoch, as opposed to the neural signal whose phase is not correlated with that of the scanner acquisition. Temporal alignment of the artifact means that there is a direction in data space where all the artifacts from the different epochs can sum up to a very large signal, while the un-aligned neural activity cancels out. Therefore the principal components corresponding to the largest eigenvalues will capture the artifact. Figure 8.8D shows examples of the first three principal components reflecting the artifact. After projecting out the artifactual components we concatenate the epochs to recover a continuous time series (fig. 8.8F). The effect of the cleaning procedure becomes evident when looking at stimulus induced activity: Figure 8.8B shows the same data as in A after cleaning. Neural activity in the stimulus on-periods is clearly stronger than in the stimulus-off periods.

8.6 Applications of tkCCA to real data

This section summarizes applications of tkCCA to multimodal data for the analysis of neurovascular coupling mechanisms. Appendix B lists the experimental details of the setup. In short, neural activity was measured simultaneously with intracranial electrophysiological recordings and high resolution fMRI in primary visual cortex. Evoked activity was recorded during a classical block stimulation paradigm. The visual stimulus was a rotating checkerboard covering the full visual field, presented for 6 seconds followed by a blank screen presented for 24 seconds. Spontaneous activity was acquired during complete darkness (for details see [Logothetis et al., 2009]). For each session we computed neural spectrograms $x(f,t)$ and voxel time series $y(s,t)$ in a spherical voxel volume around the recording electrode. Prior to the voxel extraction the spherical region of interest (ROI) was masked with an anatomical mask discarding non-brain structure. Two separate analyses were conducted, corresponding to the two complementary views of eq. (8.3) and eq. (8.4). In a first step we used tkCCA to estimate a convolution $w_x(\tau)$ from multivariate spectrograms of neural activity to high dimensional fMRI image time series (see fig. 8.2A). This allows to investigate the time frequency dynamics of the neurovascular coupling process, and the spatial dynamics in fMRI voxel space. In a second step we estimated spatiotemporal deconvolutions $w_y(\tau)$ of fMRI image time series to predict band power modulations of neural activity (see fig. 8.2B). This approach enables us to (a) estimate spatiotemporally non-separable neurovascular coupling dynamics and (b) test whether the spatiotemporal variability of the coupling dynamics contain information about neural activity.

[4]After rejection of the seven strongest components, the neural frequency spectrum did not show artifactual peaks and followed a 1/f spectrum.

Optimizing preprocessing parameters

Spatial smoothing has become a standard preprocessing operation for any fMRI analysis. The correct choice of smoothing parameters is crucial especially in the following analyses. If high frequency multivariate spatial structure contains information about neural processes, it would be disadvantageous to smooth this structure away. On the other hand, if a spatiotemporal deconvolution of over 10,000 voxels and 20 seconds at a 1Hz temporal resolution is estimated, the robustness of the estimates profits from smoothing as the danger of overfitting to noise fluctuations is decreased. Temporal and smoothing parameters were optimized using leave-one-out cross-validation. For each smoothing parameter setting we computed tkCCA on a training data set (all sessions of an experiment except the test data session) and computed the canonical correlation between estimated and true neural activity on a test data set that was not used for estimating the hemodynamic response patterns. Smoothing in time as well as smoothing in space was implemented by a parzen function that weighted each sample according to the distance of each voxel s_i to a center voxel s_0 or the distance between the time point t_i and the temporal window center t_0. The parzen function $\text{parzen}(d, \omega)$ used here is also referred to as *de la Valle-Poussin* window and is defined as

Preprocessing parameters are optimized using cross-validation

$$\text{parzen}(d,\omega) = \begin{cases} 2(1-\frac{d}{\omega})^3 & \text{if } d > \frac{\omega}{2} \\ 1-6(\frac{d}{\omega})^2+6(\frac{d}{\omega})^3 & \text{if } d \leq \frac{\omega}{2} \end{cases} \tag{8.13}$$

for a distance $d = |d_i - d_0|$ (d_i being the ith voxel or time point) from the window center d_0 and a window size ω. The parameters tested are listed in table 8.3 for fMRI data with a temporal resolution of 1Hz and 1mm^3 voxel size. Spatiotemporal smoothing can be interpreted as imposing a prior on the tkCCA solution: After filtering away the high frequency structure in time and/or space, the tkCCA estimation procedure will not be distracted by these high frequency structures and only find dependency dynamics below the spatiotemporal cutoff frequencies. Figure 8.9A/B lists the results for prediction of fMRI activity from neural data and fig. 8.9C/D shows results for prediction of neural activity from fMRI activity. The canonical correlations on test data for prediction of fMRI from neural signals is highest for temporal lowpass filtering at 0.2-0.1Hz and spatial smoothing up to a smoothing kernel width of 2mm – larger smoothing kernels result in worse performance. The canonical correlation (25th-50th-75th percentile) for the best setting during spontaneous activity

Spatiotemporal smoothing of the data imposes a prior on the deconvolution estimates

Parameter	Value
Temporal smoothing (Hz)	[0 0.3 0.2 0.1 0.05]
Temporal smoothing (seconds)	[0 3 5 9 15 19]
Spatial smoothing (mm)	[0 2 4 6 8]

Table 8.3: Smoothing parameters used for spatial and temporal smoothing

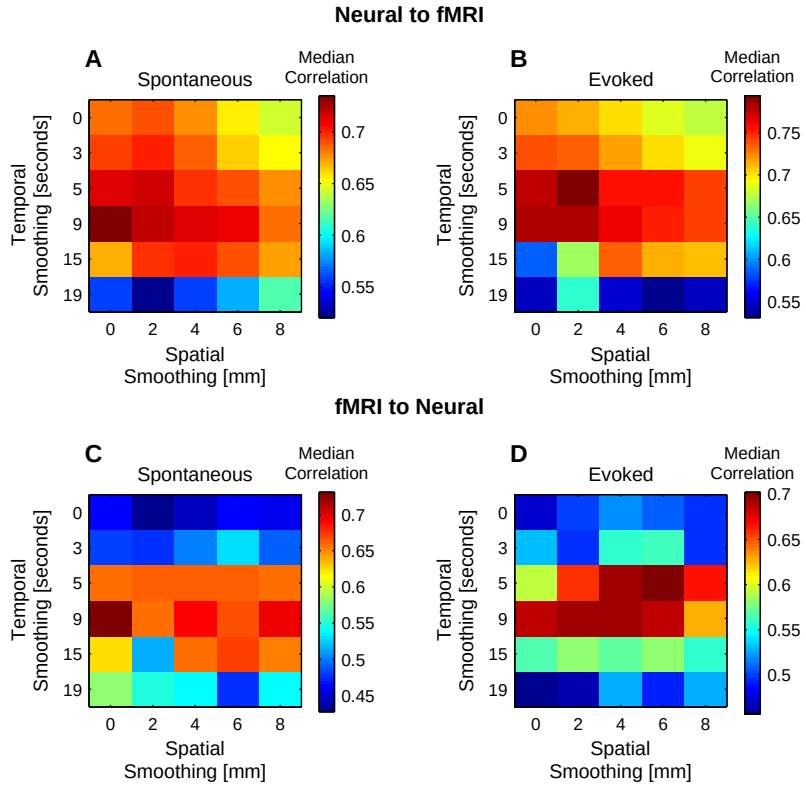

Figure 8.9: Median canonical correlation on test data across all 7 experiments for different smoothing kernel widths; **A:** Predicting fMRI signals from neural spectrograms during spontaneous activity; **B:** Predicting fMRI signals from neural spectrograms during evoked activity; **C:** Predicting neural spectrograms from voxel volume time series during spontaneous activity; **D:** Predicting neural spectrograms from fMRI during evoked activity;

was 0.71-0.72-0.76 and for evoked activity 0.76-0.79-0.81. For prediction of neural signals from multivariate fMRI time series the optimal temporal smoothing parameters are the same; however optimal spatial smoothing widths for spontaneous and evoked activity differ for this setting. Spontaneous activity is better predicted when no spatial smoothing is applied, see fig. 8.9C; fig. 8.9D shows that for evoked activity there is no clear effect of spatial smoothing on neural prediction performance. This might be due to the experimental paradigm. The visual stimulus applied activated the entire early visual cortex. Since almost every voxel carries the same information about the stimulus in this unspecific stimulation to the full visual field, the fine detail of the hemodynamic response is not important for predicting neural activity. With the optimal smoothing parameters, the correlations (25th-50th-75th percentile) be-

tween true and predicted neural activity were 0.67-0.73-0.74 (no spatial smoothing) for spontaneous and 0.65-0.70-0.71 (6mm spatial smoothing kernel) for evoked activity.

Implications for multivariate decoding analyses Leaving aside the fact that all subsequent analysis profit from a proper parameter selection for preprocessing operations such as smoothing, the optimal smoothing parameters have some interesting implications. We find that prediction of neural signals during spontaneous activity is generally worse when fMRI data was spatially smoothed (see fig. 8.9A-C). Note that when predicting fMRI signals from neural band power (fig. 8.9A/B), the drop in prediction accuracy for no or little temporal smoothing is much smaller than when predicting neural signals from fMRI (fig. 8.9C/D). Temporal high frequency structure in the fMRI signal does not seem to help for predicting neural signals. Notably increasing spatial smoothing kernel widths have different effects on the prediction of spontaneous and evoked neural activity (fig. 8.9C/D). The implications for standard fMRI analyses are that when the BOLD activity pattern is expected to have a fine scale structure, strong spatial smoothing is detrimental but temporal smoothing is necessary.

Spatial smoothing decreases prediction accuracy of fMRI signals from neural signals

Another point of this parameter selection procedure is worth noting: The results also show that the BOLD signal is predicted better from the past neural activity (corresponding to the setting in fig. 8.2A) than the neural activity is predicted from the fMRI signal (corresponding to the setting in fig. 8.2B). This is expected as the origin of fMRI signals is neural activation, and not the other way around.

Neural signals predict fMRI signals better than fMRI signals predict neural activity

Translating neural spectrograms to fMRI signals

In the following the results from the prediction of fMRI signals using time windows of neural spectrograms (see fig. 8.2A) for the best parameters (0.2Hz temporal smoothing, 2mm spatial smoothing, see fig. 8.9A/B) are presented in detail. Like classical fMRI analyses this analysis does not account for non-separable spatiotemporal dynamics of the hemodynamic response; the hemodynamic response pattern is only modeled as a temporally invariant projection w_y. But this setting can give useful insights in the time-frequency coupling of neural spectrograms. It was clear from the first studies investigating the neural basis of the BOLD signal that some frequencies of neural oscillations are reflected in the fMRI signal better than others [Logothetis et al., 2001]. What was not tested in these studies was whether there were also frequency dependent differences in the temporal dynamics of hemodynamic response. For instance higher frequencies of neural oscillations could show faster temporal neurovascular coupling dynamics than lower frequencies. This question was addressed in [Bießmann et al., 2010b, Murayama et al., 2010]. We estimated the time-frequency neurovascular coupling dynamics using tkCCA for evoked activity [Bießmann et al., 2010b] and spontaneous activity [Murayama et al., 2010].

BOLD signals reflect some neural frequencies better than others

A representative example of the data before and after tkCCA is plotted in figure 8.11. Panel A shows the mean activity of all bands of neural bandpower in blue and the mean of all voxels in red. The neural activity convolved with a standard HRF, as depicted in fig. 4.1, is plotted as purple dashed line. The purple line is temporally aligned with the red fMRI signal. The correlation coefficient between fMRI activity predicted from the canonical HRF and true fMRI activity is 0.7. In fig. 8.11B the two modalities after tkCCA are depicted. Note that also tkCCA temporally aligned the two signals. In addition to that the canonical components computed by tkCCA show a higher correlation coefficient of 0.82 between predicted fMRI signal and true fMRI activity.
Examples of the estimated time-frequency convolutions $w_x(\tau)$ are shown in figure 8.10A/B, averaged across all sessions (56 sessions for spontaneous activity and 52 sessions for evoked activity). For an easier interpretation of the temporal dimension, the time lag axis is labeled with positive time lags, corresponding to the shift in time of the BOLD signal relative to the neural signals. On top of the respective axis, denoting neural frequency bands or time lags, respectively, the strongest frequency and time lag factors are plotted in grey. Factorization was done using singular value decomposition (SVD), see e.g. [Strang, 2009]. For both stimulus conditions tkCCA robustly estimates the time frequency coupling dynamics, characterized by a pronounced peak at a time lag of about 5s in the higher frequency regime starting from 40Hz. Frequencies below 12Hz contribute less to the BOLD response. The filter obtained from evoked activity shows a slower decay and a slightly smaller amplitude at the peak response. These effects are likely to be due to the rather long duration of the visual stimulation. The 6s long block of cortical activation leads to a slower decay of the HRF. As each convolution is normalized with its vector norm this effect also leads to the slightly smaller amplitude of the peak at τ=5s. It is noteworthy that the main characteristics of the time frequency response, time point of the peak response and weighting of different frequencies, are very similar when comparing results from evoked and spontaneous activity. This can be interpreted in two ways. For one it means that the tkCCA results are not biased by stimulus induced spatiotemporal autocorrelations with respect to the main parameters of the hemodynamic coupling [Bießmann et al., 2010b]. Secondly it confirms that even from spontaneous data tkCCA can robustly estimate neurovascular coupling dynamics [Murayama et al., 2010].

Translating fMRI signals to neural spectrograms

While the coupling dynamics of single frequencies of neural oscillations are an important aspect for understanding neurovascular coupling mechanisms, the usual perspective on fMRI data is the other way around: Researchers measure fMRI activity without simultaneous acquisition of intracranial neural signals. Instead one takes an experimentally controlled stimulus time course as a proxy for the neural activity

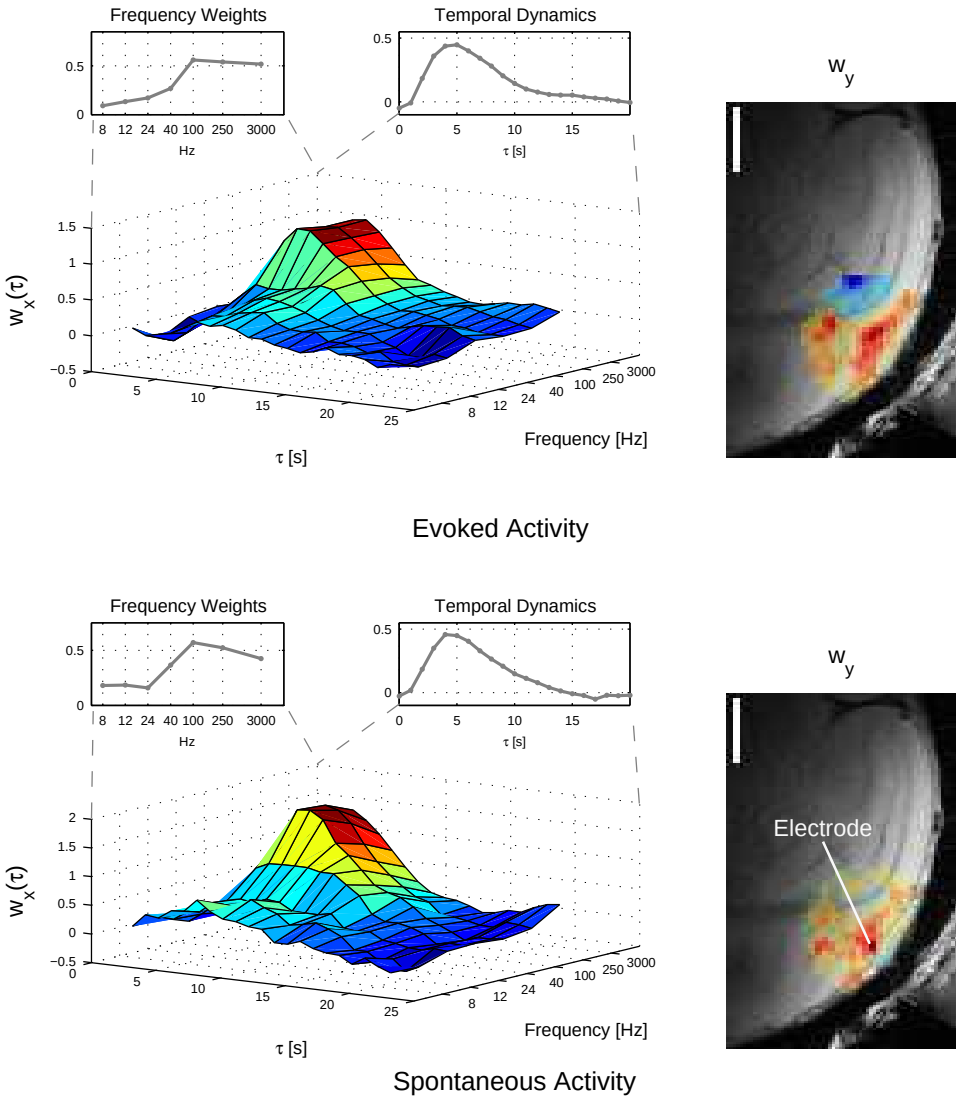

Figure 8.10: *Left column:* Time-frequency coupling dynamics $w_x(\tau)$ between neural spectrograms of different frequencies of neural oscillations and fMRI signals averaged across 7 subjects during evoked activity (*top*) and spontaneous activity (*bottom*); dominating frequency weights and temporal dynamics are plotted on top; note that evoked and spontaneous activity yield very similar coupling dynamics – consistent with most models of the HRF there is a pronounced peak at τ=5s; *right column* shows a single experiment voxel pattern w_y; scale bar in top left corner is 10-by-1mm;

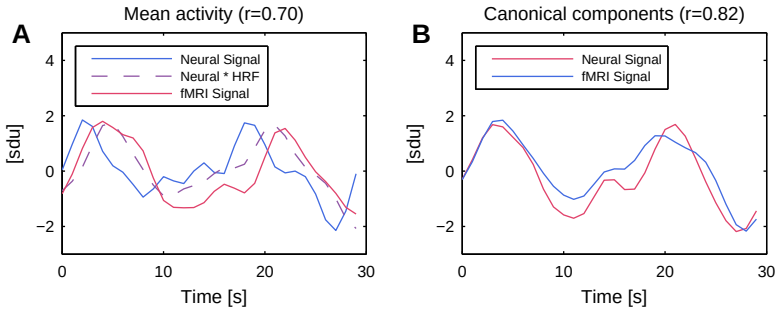

Figure 8.11: **A**: Examples of mean fMRI signal within spherical ROI around electrode (*red*) and mean neural signal (*blue*) in *standard deviation units*; note the time delay (hemodynamic lag) of about 4 seconds; convolution with HRF predicts an fMRI signal (purple dashed line) which is temporally aligned with the true fMRI signal, correlation coefficient is 0.7; **B**: Canonical components computed by tkCCA from the same data; canonical component for neurophysiological spectrograms (*blue*) and fMRI signals (*red*); tkCCA temporally aligns fMRI and neural signals without the assumption of a canonical HRF as in panel *A*; note that tkCCA yields a substantial increase in prediction accuracy of neural signals, indicated by a correlation coefficient of 0.82;

of interest. The goal of most fMRI analyses is to infer the neural activity from multivariate voxel time series. In the following we will show how to use tkCCA in order to estimate a spatiotemporal filter that translates fMRI activity into neural spectrograms. We use this deconvolution to decode neural activity from time series of voxel volumes. The simultaneous recordings allow to draw direct conclusions about the relationship between fMRI signals and neural activity.

Spatiotemporal dynamics of the BOLD signal When a neural population is activated, the hemodynamic response is not instantaneous. The HRF exhibits complex spatiotemporal dynamics, which are mainly governed by the structure of the vascular system (see fig. 3.2). Many fMRI analysis approaches assume that the spatiotemporal dynamics of the HRF is separable into a spatial and a temporal component. The spatiotemporal separability assumption is a convenient approximation. It implies that the temporal dynamics of the HRF are the same for each voxel. The function used for modeling this response is hence often called *canonical* HRF. This canonical HRF is used to temporally align a regressor, such as an experimentally controlled stimulus, with the fMRI signal, see fig. 8.11. The spatial aspects of hemodynamic activation are typically estimated using mass-univariate approaches [Friston et al., 1994, 1995, Worsley and Friston, 1995, Friston and Buchel, 2003] or multivariate pattern analysis [Cox and Savoy, 2003, Haxby et al., 2001, Haynes and Rees, 2005, Kriegeskorte et al., 2006]. There is empirical evidence justifying this approximation. Spatiotemporal separability of the hemodynamic response has

For faster computations most fMRI analyses assume spatiotemporal separability of the HRF

been reported in studies on non-human primates [Sirotin et al., 2009] and humans [Shmuel et al., 2007]. These studies however used rather simple models of the spatial profile of the hemodynamic response [Sirotin et al., 2009]. And although the results of [Shmuel et al., 2007] show that large parts of the hemodynamic response can be accounted for by a separable approximation, the authors also mention that some spatiotemporal variability of the HRF cannot be explained by separable models. Indeed many other studies provided empirical evidence showing that the HRF varies between brain regions, subjects [Aguirre et al., 1998, Handwerker et al., 2004] and cortical depth [Yacoub et al., 2006]. In particular the local fine structure of cortical vascularization gives rise to spatiotemporal hemodynamics that cannot be captured by a combination of an independent spatial and temporal component [Berwick et al., 2008]. Moreover it can be shown that this spatiotemporal variability changes depending on the length and intensity of the sensory stimulus used [Logothetis, 2003, Berwick et al., 2008].

Taken together, while the separability assumption of the HRF is convenient, several studies provide evidence against it. Consequently non-separable models can reveal areas of activation that separable models would fail to detect, see for instance [Woolrich et al., 2004, Lu et al., 2007, Mourão-Miranda et al., 2007, van Gerven et al., 2010]. Many fMRI decoding studies have the goal to decode neural activity from fMRI signals. If the spatiotemporal variability of the HRF contains information about neural processes, then accounting for non-separable hemodynamics will increase the decoding accuracy of neural signals from fMRI data. Although this hypothesis is highly relevant for many fMRI studies, it was not tested directly so far. Using tkCCA to estimate non-separable spatiotemporal deconvolutions from simultaneous recordings of neurophysiology and fMRI signals, it can be shown that non-separable spatiotemporal hemodynamics in response to neural activation indeed contain information about the underlying neural activity [Bießmann et al., 2011a]. This will be explained in more detail in the following.

High gamma band contributes most to fMRI signal Since the neural band power time series was modeled as a linear combination w_x of frequency bands, see equation (8.4), it is important to ensure that the frequency combinations found by tkCCA are physiologically plausible. Figure 8.12C shows the grand mean and standard error of mean of the normalized band power coefficients across all 7 experiments. The coefficients for each band were similar across experiments and were in line with previous findings [Logothetis et al., 2001, Niessing et al., 2005, Goense and Logothetis, 2008]: Lower frequencies contributed relatively little to the BOLD signal and the band with the largest coefficients was the high gamma band (60-100Hz).

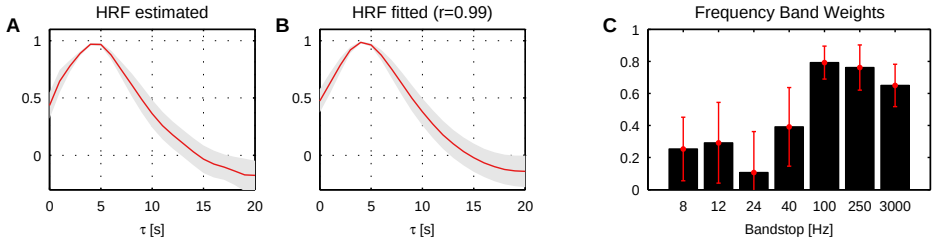

Figure 8.12: **A**: Temporal component w_τ (mean $\pm$ standard error of mean across subjects) exhibit typical characteristics of commonly used HRF models, a peak at $\tau \approx 5$s and a late undershoot; **B**: Fits of canonical HRFs using standard HRF model (as implemented in SPM5); correlation coefficient r between estimated and fitted HRFs was 0.99; **C**: Normalized frequency coupling coefficients $\alpha(f)$, averaged across all experiments (n=7); neural oscillations in the range from 60-100Hz contribute most to the BOLD signal, lower frequencies and high-frequency bands contributed less; note the similarity of the frequency coupling parameters with those obtained in the time-frequency convolution in fig. 8.10

Decomposition of spatial and temporal dynamics When comparing tkCCA with classical analysis approaches it is important to ensure that the temporal dynamics w_τ used as canonical HRF for the classical analyses is correct. We tested whether the temporal dynamics in w_τ are well captured by the commonly assumed canonical HRF model (see fig. 4.1). A popular HRF model, as implemented in the standard fMRI analysis package SPM [Friston and Buchel, 2003], was fitted to the mean w_τ for each experimental subject. This model explained w_τ very well; the grand average, across all 7 experiments, of the mean squared error between normalized fits and tkCCA estimates of w_τ was $8.8 \cdot 10^{-5}$, the correlation between the the SPM HRF model and the tkCCA estimate was 0.99. Examples for w_τ and the canonical HRF fits are depicted in fig. 8.12A/B, respectively.

Model comparisons A comparison of the spatiotemporal dynamics obtained with tkCCA and classical analysis methods is depicted in fig. 8.13. The left panel shows the temporal dynamics w_τ. Vertical lines in the temporal dynamics plot indicate the time lags at which the spatial patterns in the panels on the right side have been extracted. The color of the frame corresponds to the color of the vertical line in the left panels. Figure 8.13A shows that the non-separable tkCCA model learned a non-separable spatiotemporal pattern: Some voxels follow the 'main hemodynamic response', reflected in the factorized temporal dynamics w_τ. An example are the voxels close to the electrode, indicated as 'peak response voxels'. Other voxels show different temporal dynamics. For instance some voxels, indicated as 'undershoot voxels', show a stronger decrease in the undershoot phase of the HRF. Figures 8.13B/C show deconvolutions obtained with multivariate and mass-univariate separable models.

The spatial dynamics of these separable models are – by design – the same for all time lags, except a voxel-independent scaling factor given by w_τ.

Neural signal prediction accuracy as function of ROI radius Figure 8.14A lists the correlations (median across all sessions) obtained with the non-separable deconvolutions and all separable deconvolutions for spontaneous activity and fig. 8.14B for evoked activity. At ROI radii of 4-6mm at least every second session achieves correlations above 0.5, independent of the analysis method used. All models predicted neural activity worst for the smallest radius of 2mm (see fig. 8.14A/B). While mass-univariate deconvolutions and spatial averaging do not show increased prediction performance beyond of radii of 4mm, the predictions of the two multivariate methods $w_y^*(\tau)$ and $w_y^{mv}(\tau)$ saturate slightly later at ROI radii of 6mm.

Non-separable spatiotemporal hemodynamics contain neural information If non-separable spatiotemporal variability of the HRF contains neural information then tkCCA will predict neural signals better than separable models. In contrast if the true HRF were separable in space and time, then the predictions of the separable models would be greater or equal to those obtained with tkCCA. Figure 8.14C/D shows the percentage of sessions in which tkCCA predicted neural activity better than separable models. For a ROI size of up to 4mm there is no significant difference between the separable and non-separable multivariate model. For ROI sizes larger than 4mm, tkCCA consistently predicts neural signals better than any of the separable models. Red symbols in fig. 8.14C/D indicate that the medians of the correlations obtained with tkCCA are significantly larger than those of the separable models ($p<0.01$, Bonferroni corrected non-parametric sign test [Hollander and Wolfe, 1999]) for a given ROI size. The separable model $w_y^{mv}(\tau)$ whose predictions were closest to those of the non-separable model is directly compared to the tkCCA result in the scatterplots of fig. 8.15. Each dot is one session. Dots falling above the red line indicate that tkCCA is better than the separable multivariate model. Light grey dots indicate that this effect is not significant for a given ROI radius. Non-separable deconvolutions predict neural activity consistently better for ROI sizes larger than 4mm.

Differences between non-separable and separable models in space and time In order to examine which spatial and temporal features gave rise to this superiority of the non-separable model we compared the deconvolutions of the separable and the non-separable multivariate model. For each session we partitioned the deconvolution $w_y^*(\tau)$ and $w_y^{mv}(\tau)$ into time slices and spherical shells around the electrode.

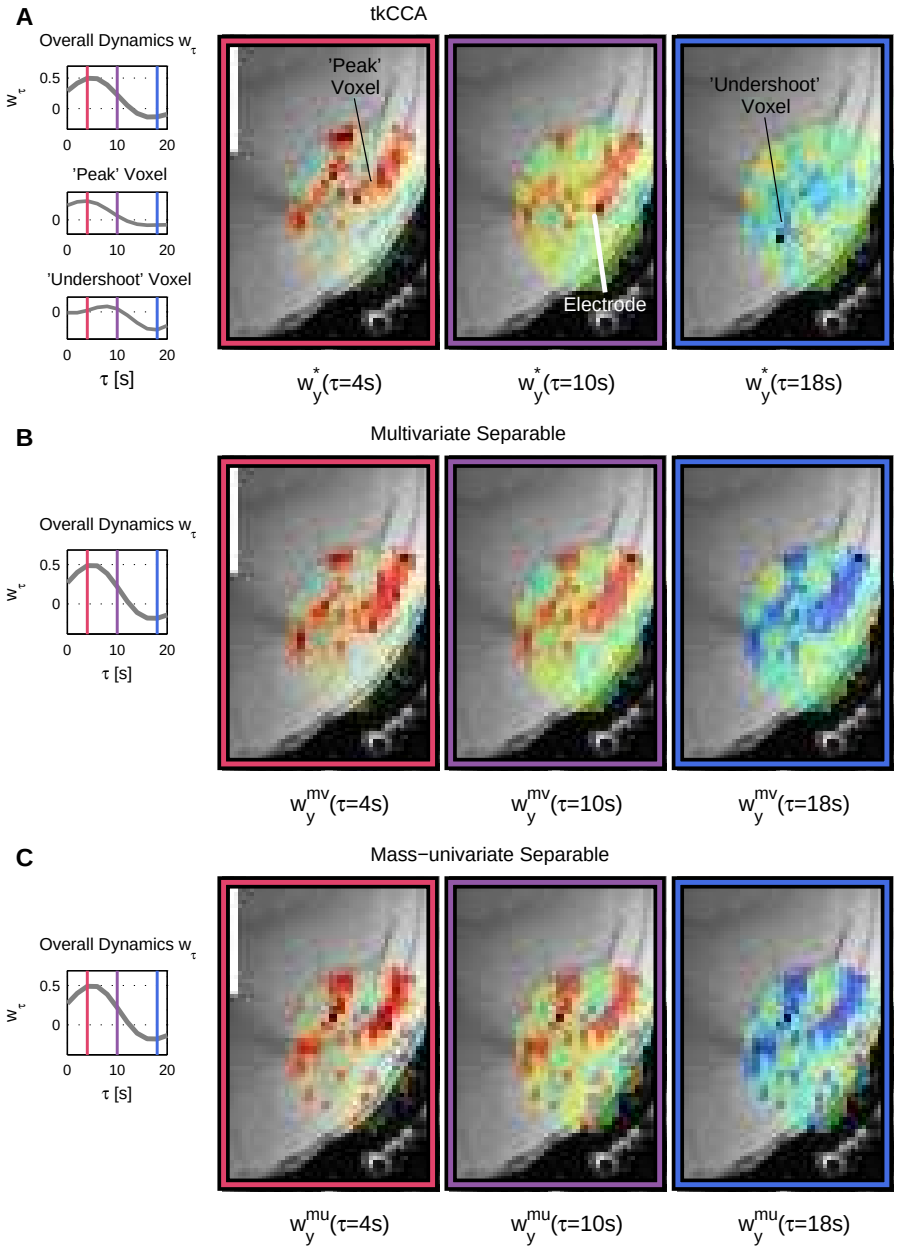

Figure 8.13: Examples of spatiotemporal deconvolutions $w_y(\tau)$ for three time lags τ plotted on transversal section of V1 (frame color corresponds to vertical lines in left panels); **A**: Non-separable tkCCA model $w_y^*(\tau)$ shows different temporal dynamics in different voxels; *left column* shows overall temporal dynamics w_τ (*top panel*), temporal dynamics in a 'peak voxel' (*middle*) and in 'undershoot voxel' (*bottom*); **B**: Separable multivariate model $w_y^{mv}(\tau)$ shows the same temporal dynamics in *all voxels*; **C**: Separable mass-univariate model $w_y^{mv}(\tau)$ also exhibits the same temporal dynamics in all voxels;

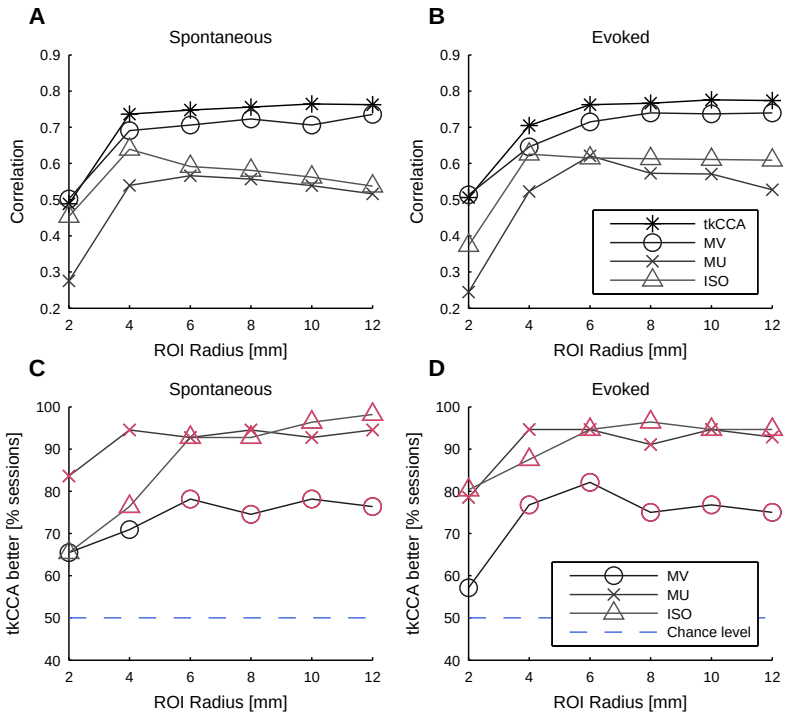

Figure 8.14: Comparison of correlations between true and estimated neural signals, obtained with non-separable (tkCCA) and separable deconvolutions for different ROI sizes; **A**: Correlation between predicted and true neural signals during spontaneous activity; **B**: Same as *A* but for evoked activity **C**: Percentage of sessions in which non-separable models predicted neural signals better than respective separable model; dashed blue line indicates random difference in performance; red symbols denote significant difference at p<0.01 (Bonferroni corrected); **D**: Same as C for evoked activity;

The time slices are intervals of lengths 1s and can be defined as sets of time delays τ that have an absolute distance of less or equal than half a second to the time slice index $t = \{1, 2, \dots\}$

$$T_t = \left\{ \tau, \ \left| \tau - t \right| \leq \frac{1}{2} \right\} \tag{8.14}$$

The spherical shells are concentric around the electrode and have a thickness of 1mm. They can be parametrized by a radius parameter $r = \{2, 3, \dots\}$ and are defined by the set of all voxels whose distance from the electrode lies in the interval $[r - \frac{1}{2}, r + \frac{1}{2}]$ mm.

$$S_r = \left\{ s, \ \left| \operatorname{dist}(s, s_0) - r \right| \leq \frac{1}{2} \right\} \tag{8.15}$$

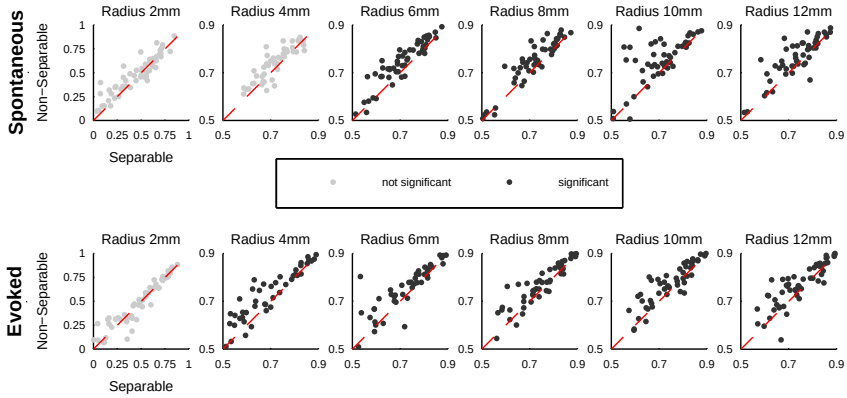

Figure 8.15: Correlations between true and neural activity estimated by the best separable devonvolution $w_y^{mv}(\tau)$ and the non-separable (tkCCA) deconvolution $w_y^*(\tau)$; tkCCA predicts neural activity better for ROI radii larger than 4mm; dashed red line indicates equal performance, each dot a session; dots falling above the diagonal indicate tkCCA is better than the best separable model; light grey indicates the medians of separable and non-separable correlations are not significantly different at $p < 0.01$ (Bonferroni corrected);

where $\text{dist}(s, s_0)$ denotes the euclidean distance between the voxels with the index s and the electrode voxel s_0. For each time slice T_t and each spherical shell S_r we calculated the squared difference between the optimal non-separable deconvolution $w_y^*(s, \tau)$ (estimated by tkCCA) and the filter obtained by separable models:

$$\left(w_y^*(s,\tau) - w_y^{mv}(s,\tau)\right)^2, \quad \forall s \in S_r, \forall \tau \in T_t \tag{8.16}$$

Figure 8.16 shows spatiotemporal difference profiles comparing separable and non-separable hemodynamic response models during evoked and spontaneous activity. Plotted are the averages over all 7 experiments. The multivariate separable and non-separable deconvolutions show similar spatial dynamics close to the electrode during the initial main response from τ=4-8s. With increasing cortical distance the patterns become more dissimilar. In a time window from τ=10s to τ=16s the two approaches yield very different spatial patterns, especially at voxels far away from the recording position. This analysis shows that non-separable spatiotemporal dynamics are found mainly in distant voxels in the later phase of the HRF.

Implications for fMRI analysis methods

Isotropic spatial smoothing predicts neural signals worst The fact that isotropic spatial averaging was consistently worse than the non-separable deconvolutions suggests that the spatial structure of the optimal spatiotemporal deconvolution is not a

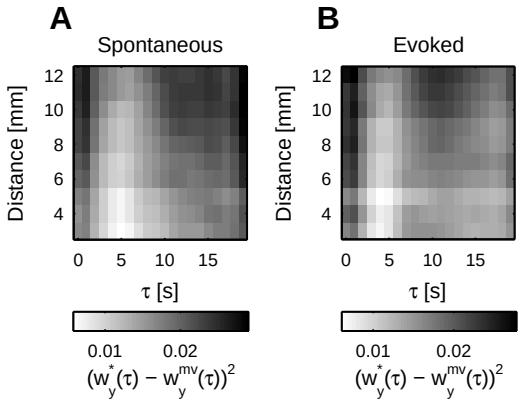

Figure 8.16: Difference between non-separable and separable multivariate deconvolutions as function of time lag and cortical distance during *spontaneous* (A) and *evoked* (B) activity averaged across all experiments; note the high similarity of non-separable and separable deconvolutions close to the electrode at time lags of about 5s when the HRF is peaking; with cortical distance the two models become less similar, indicating non-separable dynamics; distant voxels at time lags around 10s exhibit most non-separable features;

spatially isotropic filter in combination with a canonical HRF. A standard preprocessing step in fMRI analyses is to spatially smooth the fMRI data and to construct a regressor for each voxel by temporally smoothing the experimentally controlled variables with a canonical HRF. This preprocessing step corresponds to the isotropic smoothing setting. Our results show that accounting for non-separable spatiotemporal dynamics improves predictions of neural activity. This makes sense, considering the anisotropic structure of primary visual cortex: Along the cortical laminae there are different correlational structures than perpendicular to the cortical laminae. Isotropic spatial smoothing does not take this into account. This finding can be transferred to standard fMRI experiments: Consistent with recent findings on multivariate decoding of fMRI activity [Chen et al., 2011] our results show that using anatomically or functionally informed priors leads to better results in subsequent analyses.

Multivariate vs. mass-univariate methods The underlying assumption of mass-univariate analyses is that the hemodynamic response to neural activation can be approximated by analyzing time series of single voxels separately. We tested this spatial independence assumption by predicting neural activity from mass-univariate correlation estimates of the spatial hemodynamic response pattern. Indeed the mass-univariate approach predicted neural activity reliably with correlations above 0.5. Compared to multivariate estimates however mass-univariate predictions were consistently worse (see fig. 8.14). This effect is consistent with theoretical studies such

as [Kriegeskorte et al., 2006] comparing mass-univariate and multivariate estimates of neural activity from multivariate voxel time series. The superiority of multivariate decoding is unlikely to be due to wrong estimates of the temporal dynamics w_τ obtained from the non-separable model $w_y^*(\tau)$. As shown in fig. 8.12A/B, w_τ obtained from factorizing the tkCCA solution is fitted well by a standard canonical HRF model.

Optimal searchlight radii for decoding neural signals Our results quantify how much information about neural activity is lost when choosing wrong radii in standard searchlight approaches [Kriegeskorte et al., 2006]. We found lower decoding accuracies for searchlight radii smaller than 4mm indicating that there is information about local neural activity in the multivariate structure around the electrode that is missing in small ROIs. This information is not captured by spatial averaging nor mass-univariate analyses. On the other hand ROIs larger than 6mm do not increase the prediction accuracy, suggesting that at distances more than 6mm away from the site of neural activation, there is mostly redundant information about local neural activity in the fMRI signal.

Non-separable vs. separable deconvolutions Space-time separability of the hemodynamic response to neural activation has been reported previously in optical imaging studies on rat barrel cortex for the oxyhemoglobin and the total hemoglobin response [Devor et al., 2005]; the spatiotemporal dynamics of deoxyhemoglobin[5] seems to have a more complex structure. All three hemodynamic signals however showed very similar spatial patterns at the initial peak of the HRF in this study. Optical imaging studies in alert monkeys refined this result by showing that the *initial dip* of the rapid deoxyhemoglobin response had the same spatial dynamics as the slower blood volume response [Sirotin et al., 2009], also speaking in favor of spatiotemporal separability. However the authors of [Sirotin et al., 2009] assume an isotropic gaussian profile for the spatial dynamics and consider only the estimated width parameter. Such a simple spatial profile seems at odds with the complex cortical microvasculature [Weber et al., 2008] (see also fig. 3.2), other optical imaging results [Berwick et al., 2008] and our finding that anisotropic filters predict neural activity significantly better than isotropic gaussian profiles. It is worth keeping in mind that while the results of optical imaging studies offer a very high in-plane resolution, they are limited to superficial cortical layers down to $\approx 400\mu$m. There is evidence from high resolution fMRI studies showing that cerebral blood volume (CBV) changes in the surface vessels return faster to baseline levels after stimulation than CBV changes in deep cortical layers [Yacoub et al., 2006]. Also in humans the response to highly localized visual stimuli seems to be well approximated by separable dynamics: In [Shmuel et al., 2007] it was shown that the the first spatial

[5]Deoxyhemoglobin is an important factor contributing to the BOLD signal [Ogawa et al., 1992]

principal component of the hemodynamic point spread function accounts for 96% of the variance in the fMRI signal. The authors also report variability of the point spread function over time. Our results confirm these results in that separable approximations are very similar to non-separable hemodynamic response patterns – especially in an early time window directly after a neural impulse (see fig. 8.16). We found however that for decoding neural activity, non-separable models are consistently better than separable models. Thus the variability of the hemodynamic response that is not explained by separable models contains information about neural activity. This is consistent with findings from studies without neural measurements. In [Mourão-Miranda et al., 2007] the authors showed that taking into account the spatiotemporal dynamics of the hemodynamic response to a stimulus variable can help to find patterns of activation that are difficult to detect with classical methods.

Limitations of the decoding approach Our results show that all methods tested yield accurate predictions of neural activity. However non-separable deconvolutions predicted neural activity consistently better on test data confirming that non-separable dynamics of the hemodynamic response contain information about neural activity. For the correct interpretation of our results it is important to keep in mind that we cannot rule out that systematic time delayed correlations between the neural activity recorded and neural activity within all other voxels distort our estimate of spatiotemporal filters. An example of such a confound would be slowly traveling waves of neural activity [Leopold et al., 2003]. While this potential confound might bias our deconvolution estimates, there is evidence from optical imaging studies showing that the spatiotemporal hemodynamic response several millimeters away from the ”center voxel” is not reflecting time delayed neural correlations but really a hemodynamic effect [Devor et al., 2005]. Whether or not distant neural activity is affecting our estimates, from a decoding point of view these potential confounds are neglegible. The aim of many fMRI analyses is not the exact estimation of the hemodynamic response pattern but rather an accurate prediction of neural activity. Moreover if the scientific hypothesis can be formulated in a decoding framework, then the only quantity of interest for falsifying the hypothesis is the accuracy with which neural activity can be predicted from fMRI signals. Our results show that tkCCA can be used to predict neural activity better than standard methods by exploiting spatiotemporal non-separable hemodynamics. We conclude that including non-separable spatiotemporal dynamics in the analysis of fMRI data will help to improve inference about neural activity from hemodynamic signals, especially if the hemodynamic data has a high spatial resolution.

Chapter 9

Summary and outlook

Technical advances enable researchers to measure brain activity at an increasingly high level of detail. The usefulness of these advances depends critically on the choice of an appropriate analysis method. Over the years, many methods have been proposed, from biologically realistic models as in [Daunizeau et al., 2007, Riera et al., 2007, Valdes-Sosa et al., 2009, Bojak et al., 2011], over multivariate data driven approaches [Cox and Savoy, 2003, Norman et al., 2006, Haynes and Rees, 2006] and less complicated mass-univariate approaches [Friston and Buchel, 2003] to simple analyses based on averaging and bivariate correlation coefficients [Logothetis et al., 2001, Martindale et al., 2003]. This dissertation proposes a compromise between these approaches. We introduce tkCCA as a simple but powerful data-driven algorithm for the analysis of high resolution multimodal neuroimaging data.

Biologically realistic models are necessary and useful ...

..., but not every scientific hypothesis needs biologically realistic models

Good models of neurophysiological and hemodynamic activity are indispensable for testing hypotheses about physiological processes and of great value for connecting different levels of research – intracellular recordings, large scale connectivity data and functional measurements. But the appropriate tradeoff between biological realism and tractability depends on the scientific question addressed. For instance in EEG forward modeling, a conductivity profile of the head is important if one wants to estimate the exact location of a current source. If however the goal of a study is not localization but if the scientific hypothesis can be cast in a decoding setting, then dipole fitting is not needed – after all the EEG forward model is just a linear mapping, see eq. (4.18). If the scientific hypothesis can be tested by decoding a stimulus variable from the EEG time series, any linear classifier could 'undo' the volume conduction effects for the classification, or in general the decoding task. Fitting a dipole model to the data will not increase the amount of information in the data. To summarize, not every hypothesis needs a forward model furnished with all biological details accumulated in empirical research. In fact including too much biological realism often renders it difficult to assess whether or not a particular feature is needed to explain the data.

The scientific question addressed with tkCCA was: Can we increase the amount of information about neural activity extracted from fMRI signals by abandoning the assumption that the hemodynamic response to neural activation is separable in space and time? While previous studies provided plenty of indirect evidence in favor of this hypothesis, it has not yet been investigated directly. In order to answer this question we chose a decoding approach. Using a spatiotemporal deconvolution we decoded intracortical neural activity from fMRI voxel time series. The spatiotemporal deconvolutions were constructed such that they reflected the main assumptions underlying popular fMRI analysis approaches. In particular all models except the tkCCA model assumed spatiotemporal separability of the hemodynamic response. In this setting the above hypothesis can be easily translated into different experimental outcomes: If the non-separable parts of the hemodynamic response do carry information about neural activity, then the non-separable tkCCA deconvolution will predict neural activity better than separable deconvolutions.

Our results show that the predictions of non-separable deconvolutions estimated with tkCCA were consistently better than those of separable deconvolutions, corresponding to isotropic spatial smoothing, mass-univariate analyses and multivariate decoding approaches. This represents to the best of our knowledge the first direct empirical evidence in favor of the hypothesis that the spatiotemporally non-separable part of the hemodynamic response contains information about neural activity. A biologically realistic model for the complete spatiotemporal dynamics of high resolution fMRI has not been established yet. Such a model requires basic research about the neurovascular coupling mechanisms. And analysis of these experimental data requires methods that do not rely on restrictive hypotheses. The approach taken here might seem simple, yet it yields new insights about the spatiotemporal dynamics neurovascular coupling: We can provide spatial and temporal parameters describing where non-separable dynamics occur and how important they are for decoding neural activity. These results could not have been obtained with most of the classical analysis methods. Our findings thus might help to build more detailed models of the spatiotemporal neurovascular coupling dynamics. Moreover tkCCA can of course also increase the prediction accuracy of neural signals in standard non-invasive fMRI experiments.

Directions of future research

The application examples on real neuroimaging data were restricted to one brain area only. This allowed for a detailed examination of neurovascular coupling mechanisms in an anatomically confined region. Future work will focus on application of the presented algorithm to completely non-invasive EEG-fMRI data in more than one brain area. Potential areas of application include hybrid brain computer inter-

faces such as in [Fazli et al., 2011], but also the detection and localization of epileptic activity. In contrast to the matrix data of the neurophysiology spectrograms this would involve tensor data, spanning the dimensions of frequency, time and EEG channels. As the solution of tkCCA is computed on inner products of data points, it naturally extends to tensor data.

Moreover tkCCA is ideally suited to study changes in neurovascular coupling mechanisms. In the application examples presented, the well controlled experimental conditions guaranteed stationarity of the neurovascular coupling dynamics. However, certain neuromodulators can change the hemodynamic response to neural activation. Preliminary results show that local application of Acetylcholine increases the baseline activity of the BOLD signal [Rauch et al., 2009]. Comparing the neurovascular coupling parameters estimated by tkCCA one can detect changes in the spatiotemporal dynamics of neurovascular coupling. This approach could yield new insights in neurovascular coupling pathologies associated with cholinergic dysfunction, such as Alzheimer's disease.

Another direction of future research is the exploration of non-linear kernel functions for tkCCA. As the visualization of the results is of particular importance in neuroimaging applications, we opted for linear kernels in the analyses presented. However if the visual inspection of the features is not essential but only the question whether two modalities are coupled or not, then non-linear kernels in combination with tkCCA are a powerful alternative to standard methods. Non-linear kernel functions allow to estimate arbitrary non-linear state space models [Yamada and Azimi-Sadjadi, 2005]. When applied to multimodal data, tkCCA with non-linear kernels can also quantify information flow [Wu et al., 2011] in a *Granger-Causality* sense [Granger, 1969].

Of course tkCCA can be readily applied to any data beyond the field of neuroimaging. Preliminary results have been presented in [Bießmann and Harth, 2010], in which dependency structures in high dimensional graphs extracted from a social platform on the internet were analyzed. Similar approaches can be taken for analyzing news feeds crawled from the internet. Web data are inherently multimodal as they comprise pictures, text data, geographical information, structural markers and many more features that might contain useful information. Current web data analysis frameworks do not account for the dynamic aspects of web graphs. First attempts to tackle dynamic aspects of large web data sets have been presented in [Leskovec et al., 2009]. Future work involving more efficient implementations of kCCA, for instance maximum-margin based approaches to CCA as presented in [Szedmak et al., 2007], will be helpful for tackling larger data sets than those we dealt with in the neuroimaging data applications.

Appendix

Appendix A

Mathematical preliminaries

A.1 Generalized Eigenvalue Equations

An *eigenvector* $w \in \mathbb{R}^n$ of a square matrix $A \in \mathbb{R}^{n \times n}$ is defined as a vector that is mapped to a multiple of itself when left-multiplied with A:

$$Aw = \lambda w. \tag{A.1}$$

The corresponding factor λ is called *eigenvalue* and eq. (A.1) is called an *eigenvalue equation*. The scale of the eigenvectors is undefined, since for every eigenvector w, a rescaled version αw with $\alpha \in \mathbb{R}$ is also an eigenvector with the same eigenvalue. So it is useful to think of eq. (A.1) as defining only eigendirections, i.e. directions that do not change when projected on A.

From the eigenvalue equation we see that the matrix $A - \lambda I$ has not full rank, since $(A - \lambda I)w = 0$, so we can find the eigenvalues of the matrix A by finding the zero set of the characteristic polynomial, i.e. by solving

$$\det(A - \lambda I) = 0$$

for λ. Since this is a n-th order polynomial in λ, this implies that there are always n eigenvalue/eigenvector pairs. Note however, that even for real matrices A the eigenvalues are not necessarily real, since in general even real polynomials have complex roots. If we write all the n eigenvectors as columns of a matrix W, we can write eq. (A.1) as matrix equation

$$AW = W\Lambda \tag{A.2}$$

where Λ is a diagonal matrix with the eigenvalues λ on its diagonal. Equation (A.1) is a special form of the *generalized eigenvalue equation*

$$Aw = \lambda Bw \tag{A.3}$$

with $A, B \in \mathbb{R}^{n \times n}$ or in matrix form

$$AW = BW\Lambda \tag{A.4}$$

and the generalized eigenvalues can be found by solving $\det(A - \lambda B) = 0$. Often the matrices corresponding to A and B will be real and symmetric matrices computed from measured data, for instances if A, B are covariance matrices of kernel matrices. This not only leads to real eigenvalues and eigenvectors, but also to a certain orthogonality structure of the eigenvectors. Consider two different eigenvalue/eigenvector pairs (λ_i, w_i) and (λ_j, w_j). Multiplying the respective generalized eigenvalue equations with $w_j^\top$ and $e_i^\top$ from the left, we obtain

$$w_j^\top A w_i = \lambda_i w_j^\top B w_i \qquad \text{and} \qquad w_i^\top A w_j = \lambda_j w_i^\top B w_j$$

Symmetry of A and B means that either $\lambda_i = \lambda_j$ or

$$w_j^\top A w_i = w_j^\top B w_i = 0. \tag{A.5}$$

so that eigenvectors with different eigenvalues are orthogonal with respect to the scalar products defined by A and B. For ordinary eigenvalue problems, where $B = I$, the eigenvectors are orthogonal wrt. the standard euclidean scalar product. In matrix notation this means that both matrices A and B are simultaneously diagonal with respect to the basis of the generalized eigenvectors

$$W^\top A W = D_A \qquad \text{and} \qquad W^\top B W = D_B \tag{A.6}$$

where D_A and D_B are diagonal matrices. Therefore, solving the generalized eigenvalue problem for two symmetric matrices is sometimes also called *simultaneous diagonalization*. For a detailed introduction to linear algebra see for instance [Strang, 2009]. An advantage of eigendecompositions is that – depending on how one computes the matrices A and B – one can cast a large number of algorithms in this framework and apply standard eigenvalue solvers. This allows to understand, implement and extend many existing statistical learning algorithms using only equation (A.4). Generalized eigenvalue equations are a generic way of solving multivariate regression problems. Among the many applications are e.g. t-tests, analysis of variance (ANOVA) [Henson, 2000], partial least squares, ridge regression [Bie and Moor, 2003], (kernel) fisher linear discriminant analysis [Mika et al., 1999], blind source separation [Diamantaras and Kung, 1996] or clustering [von Luxburg, 2007]. Despite the fact that eigenvalue equations are linear equations, they are not restricted to linear data analysis: Many non-linear methods can be solved efficiently in the form of equation (A.4) using non-linear kernel functions, see e.g. [Schölkopf and Smola, 2001, Schölkopf et al., 1998, Shawe-Taylor and Cristianini, 2004]. There is a simple intuition underlying eigendecompositions: one wants to maximize the variance in some subspace of the data (expressed in how one computes A in eq. (A.4)), while minimizing or constraining the variance in another subspace of the data (expressed in how one computes B in eq. (A.4)). For a concise review of eigenvalue problems and their application to unsupervised and supervised analysis, such as regression, latent variable recovery and clustering, the reader is referred to e.g. [Bie et al., 2005, Shawe-Taylor and Cristianini, 2004].

A.2 Relation between PCA and linear kernel PCA

The relation between the standard PCA on the covariance matrix $C_X = XX^\top$ and the linear kernel version on the kernel $K_X = X^\top X$ can be understood best by performing a *singular value decomposition* (SVD) (see e.g. [Strang, 2009]) of the data matrix X

$$X = ESF \tag{A.7}$$

where $E \in \mathbb{R}^{D\times N}$ and $F \in \mathbb{R}^{N\times D}$ are orthogonal matrices and S is a diagonal matrix containing the so-called singular values. It can be shown that $E \in \mathbb{R}^{D\times N}$ and $F \in \mathbb{R}^{N\times D}$ contain the eigenvectors of $XX^\top$ and $X^\top X$, respectively. The eigenvalues of $XX^\top$ and $X^\top X$ are the squared singular values on the diagonal of S. It is straightforward to see this when computing C_X and K_X from the SVD solution:

$$\begin{aligned} C_X &= XX^\top \\ &= ESF(ESF)^\top \\ &= ESFF^\top S^\top E^\top \\ &= ESIS^\top E^\top \\ &= ES^2E^\top \end{aligned} \tag{A.8}$$

and

$$\begin{aligned} K_X &= X^\top X \\ &= FSE(FSE)^\top \\ &= FSEE^\top S^\top F^\top \\ &= FSIS^\top F^\top \\ &= FS^2F^\top \end{aligned} \tag{A.9}$$

where we used $EE^\top = I = FF^\top$. So to summarize when $D < N$ (less dimensions than samples), it is more efficient to compute PCA on the covariance matrix $C_X = XX^\top$. When $D > N$ (less samples than dimensions), the dimensionality of the principal component subspace is at most $(N-1)$ – thus it is more efficient to compute PCA on the linear kernel matrix $K_X = X^\top X$.

Appendix B

Experimental protocols

B.1 Preparative surgery and experiment procedure

This study involved 7 combined electrophysiological fMRI experiments in 4 healthy monkeys (Macaca mulatta), weighting 7-10kg. All experiments and surgical procedures were approved by the local authorities (Regierungspräsidium) and were in full compliance with the guidelines of the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals. A custom-made plastic head holder and a recording chamber were implanted directly on to the skull of each animal under general anesthesia (balanced anesthesia consisting of isoflurane 1.3% and fentanyl 3μg/kg i.v. as needed). Experiments were conducted under general anesthesia. After premedication with glycopyrolate (0.01mg/kg i.m.) and ketamine (15mg/kg i.m.), animals were preoxygenated and intubated following induction with fentanyl (3μg/kg i.v.), thiopental (5mg/kg i.v.) and succinyl chloride (3mg/kg i.v.) and artificially ventilated (Servo Ventilator 900C; Siemens, Germany) maintaining an end-tidal CO2 of 33mmHg and oxygen saturation of over 95%. Anesthesia was maintained with remifentanil (0.5-2μ g/kg i.v.) and muscle relaxation was achieved with mivacurium (3-6mg/kg/h i.v.). Body temperature was kept at 38-39.5°*C*, Lactated Ringer's solution (2.5% glucose) was infused at 10mL/kg/h i.v.. Drops of 1% cyclopentolate hydrochloride were instilled into each eye to achieve mydriasis and hard contact lenses (Zeiss) with the appropriate dioptric power were used to bring the eyes to focus on the stimulus plane. Visual stimuli were presented binocularly using XGA fiber-optic projectors (Silent Vision; AVOTEC, Stuart, FL). Subject's fiber-scopes were aligned to the fovea of each eye using a modified fundus camera (RC250; Zeiss). The FOV of the scope was 30×23 degrees in visual angle. Visual stimuli were generated on-fly by an OpenGL based custom program. Stimulus timings and neuronal recordings were synchronized to image acquisition and controlled by the network connected PC which ran the custom program under QNX real time OS [Logothetis et al., 2001].

B.2 Data Acquisition

We made measurements in a vertical 4.7T scanner with a 40cm diameter bore (BioSpec 47/40v; Bruker Medical, Ettlingen, Germany) [Logothetis et al., 2001]. Customized, small radio-frequency coils (30-80mm diameter) were used to increase sensitivity over a recording chamber. For positioning a multi-channel electrode, a GEFI/FLASH sequence was used with a FOV of 96×96mm at a matrix of 256×256 and a slice thickness of 1mm. Multi-slice fMRI was carried out by using multi-shot (segmented) GE-EPI. Acquisition parameters of fMRI were: FOV of 64×64 or 96×96mm; 7 or 12 slices (2mm thickness); matrix size of 64×64 or 96×96; 4 segments; echo time (TE) of 15-20ms; repetition time (TR) of 250-500ms; flip angle of 20-40°; sweep width of 100kHz; 4-8sec dummy scanning. For anatomical reference, GEFI/FLASH was used with the same FOV and slice angle as fMRI but with double matrix size and 0.5-1mm slice thickness. All imaging data were reconstructed by using Bruker software. A small craniotomy was made in each experiment to position the MRI compatible multi-channel electrode in early visual areas, V1 and V2. The electrode was made of carbon composites as a carrier (300×300μm cross section), Pt/Ir (90/10) wires (18μm diameter; California fine wire, USA) as 10 contact sites in the brain and flattened Ag wires glued on the carbon shaft as ground and noise sensor in the recording chamber. Procedures for signal amplification and noise minimization during simultaneous recording based on current measurement are described in detail elsewhere [Logothetis et al., 2001, Oeltermann et al., 2007]. The noise interference caused by the alternating magnetic field gradients during fMRI has two major components; the far-interference that can be reduced by adjusted counter-current injection through a junction electrode (Ag wire) placed in the mouth, and the near-interference, that is not measurable since it is too close to the recording site to permit the placement of additional sensors. This component can be approximated by a linear sum of gradient magnetic field changes in the XYZ directions and minimized within the preamplifier by adjusting each XYZ factor. For our multi-channel electrode, the far-interference compensation was done by a separate device for all channels, while the near-interference compensation was done within each channel-amplifier. For data collection, 4 out of 10 channels were selected as visually responsive channels by using a computer-controlled receptive field plotting program. Amplified electrophysiological data were digitized with a 16bit AD card (PCI-6052E; National Instruments) at 20.83 kHz to collect broad band neural activities including local field potential, spiking activities and residual interference noise. For all measurements of spontaneous fluctuations, the eyes were kept open in darkness, with stimulus projectors powered off to avoid the effects of LCD flicker [Logothetis et al., 2009]. We collected spontaneous activity data and data during full field visual stimulation (blank-stimulus-blank/10-6-14s, visual stimulation with polar gratings, 100% contrast, rotating in full field, session duration 5 minutes). In total we acquired 53 sessions of evoked activity recordings and 56 sessions of spontaneous activity.

B.3 Data Preprocessing

Data analysis was done using custom-written software in MATLAB (The MathWorks, Natick, MA). Regions of interest (ROIs) were manually defined according to anatomically defined borders [Saleem et al., 2002]. After discarding the first 5 volumes, voxel time courses within each ROI were centered and normalized to standard deviation units (SDU) based on blank periods. In addition to the ROI comprising the entire primary and parts of secondary visual cortex, we masked the ROIs with isotropic distance dependent ROIs at distances of 2, 4, 6, 8, 10 and 12mm away from the recording site. For neuronal data broad-band signals were decimated by a factor of 3 and a PCA based denoising method was applied to remove residual interference (see section 8.5). After this removal of the scanner gradient induced artefacts, the signal was split into 7 frequency bands f_i (f_1=1-8, f_2=8-12, f_3=12-24, f_4=24-40, f_5=60-100, f_6=120-250, f_7=1000-3000Hz). Then the signal was converted into absolute amplitude, resampled to 250Hz after appropriate low pass filtering and normalized to standard deviation units based on blank periods. Frequency separation was based on results from information theoretical analysis [Belitski et al., 2008].

Bibliography

G. K. Aguirre, E. Zarahn, and M. D'Esposito. The variability of human, BOLD hemodynamic responses. *Neuroimage*, 8(4):360–9, 1998.

A. Aizerman, E. Braverman, and L. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25:821–837, 1964.

S. Akaho. A kernel method for canonical correlation analysis. In *Proceedings of the International Meeting of Psychometric Society*, 2001.

P. J. Allen, G. Polizzi, K. Krakow, D. Fish, and L. Lemieux. Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction. *Neuroimage*, 8(3):229–239, 1998.

P. J. Allen, O. Josephs, and R. Turner. A method for removing imaging artifact from continuous EEG recorded during functional MRI. *Neuroimage*, 12(2):230–9, 2000.

T. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, 1958.

T. Arthur and P. Nader. *In Vivo Optical Imaging of Brain Function*, chapter Intraoperative Optical Imaging. CRC Press, 2002.

D. Attwell and C. Iadecola. The neural basis of functional brain imaging signals. *Trends in Neurosciences*, 25(12):621–625, 2002.

D. Attwell and S. B. Laughlin. An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab*, 21(10):1133–45, 2001.

F. R. Bach and M. I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48, 2002.

F. R. Bach and M. I. Jordan. A probabilistic interpretation of canonical correlation analysis. *Technical Report*, 2006.

S. Baillet. Electromagnetic brain mapping. *Signal Processing Magazine, IEEE*, 18(6):14 – 30, 2001.

G. Bakir, J. Weston, and B. B. Schölkopf. Learning to find pre-images. *Advances in Neural Information Processing Systems*, 2004.

A. Belitski, A. Gretton, C. Magri, Y. Murayama, M. A. Montemurro, N. K. Logothetis, and S. Panzeri. Low-frequency local field potentials and spikes in primary visual cortex convey independent visual information. *J Neurosci*, 28(22):5696–5709, 2008.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–59, 1995.

H. Benali, M. Pélégrini-Issac, and F. Kruggel. Spatio-temporal covariance model for medical images sequences: Application to functional MRI data. *Information Processing in Medical Imaging*, 2082: 197–203, 2001.

C. M. Bennett, G. L. Wolford, and M. B. Miller. The principled control of false positives in neuroimaging. *Soc Cogn Affect Neurosci*, 4(4):417–22, 2009.

H. Berger. Über das Elektroenkephalogramm des Menschen. *Archiv für Psychatrie*, 87, 1929.

S. Bergstrand, M. B. Åberg, T. Niiniskorpi, and J. Wessberg. Towards unified analysis of EEG and fMRI: a comparison of classifiers for single-trial pattern recognition. *International Joint Conference on Biomedical Engineering Systems and Technologies - BIOSTEC*, pages 273–278, 2009.

J. Berwick, D. Johnston, M. Jones, J. Martindale, C. Martin, A. J. Kennerley, P. Redgrave, and J. E. W. Mayhew. Fine detail of neurovascular coupling revealed by spatiotemporal analysis of the hemodynamic response to single whisker stimulation in rat barrel cortex. *Journal of Neurophysiology*, 99 (2):787–98, 2008.

A. Bhattacharyya, F. Bießmann, J. Veit, R. Kretz, and G. Rainer. Functional and laminar dissociation between muscarinic and nicotinic cholinergic neuromodulation in the primary visual cortex. *Submitted*, 2011.

T. D. Bie and N. Cristianini. Kernel methods for exploratory pattern analysis: a demonstration on text data. *Springer Lecture Notes in Computer Science*, 3138:16–29, 2004.

T. D. Bie and B. D. Moor. On the regularization of canonical correlation analysis. *Int. Sympos. ICA and BSS*, 2003.

T. D. Bie, N. Cristianini, and R. Rosipal. Eigenproblems in pattern recognition. *Handbook of Geometric Computing : Applications in Pattern Recognition, Computer Vision, Neuralcomputing, and Robotics*, 2005.

F. Bießmann and A. Harth. Analysing dependency dynamics in web data. *AAAI Spring Symposium*, 2010.

F. Bießmann, A. Rauch, F. C. Meinecke, X. Zhang, G. Rainer, K.-R. Müller, and N. K. Logothetis. Investigating the relationship between pharmacological MRI and electrophysiology using canonical correlation analysis. *Forum of European Neuroscience*, 6:123, 2008.

F. Bießmann, A. Gretton, F. C. Meinecke, X. Zhang, G. Rainer, N. K. Logothetis, K.-R. Müller, and A. Rauch. Investigating neurovascular coupling using canonical correlation analysis between pharmacological MRI and electrophysiology. *BMC Neuroscience*, 10(Suppl 1):P86, 2009.

F. Bießmann, F. C. Meinecke, A. Bhattacharyya, J. Veit, R. Kretz, K.-R. Müller, and G. Rainer. Comparison of v1 receptive fields mapped with spikes and local field potentials. *Computational and systems neuroscience 2010 (COSYNE)*, 7, 2010a.

F. Bießmann, F. C. Meinecke, A. Gretton, A. Rauch, G. Rainer, N. K. Logothetis, and K.-R. Müller. Temporal kernel cca and its application in multimodal neuronal data analysis. *Machine Learning Journal*, 79(1-2):5—27, 2010b.

F. Bießmann, Y. Murayama, N. K. Logothetis, K.-R. Müller, and F. C. Meinecke. Improved decoding of neural activity from fMRI signals: Towards non-separable spatiotemporal deconvolutions. *Neuroimage (in revision)*, 2011a.

F. Bießmann, S. M. Plis, F. C. Meinecke, T. Eichele, and K.-R. Müller. Analysis of multimodal neuroimaging data. *IEEE Reviews in Biomedical Engineering (accepted)*, 2011b.

D. A. Boas, S. R. Jones, A. Devor, T. J. Huppert, and A. M. Dale. A vascular anatomical network model of the spatio-temporal response to brain activation. *Neuroimage*, 40(3):1116–29, 2008.

I. Bojak, T. F. Oostendorp, A. T. Reid, and R. Kötter. Towards a model-based integration of co-registered electroencephalography/functional magnetic resonance imaging data with realistic neural population meshes. *Philos Transact A Math Phys Eng Sci*, 369(1952):3785–801, 2011.

M. Borga. Learning multidimensional signal processing. *PhD Thesis, Dissertation No. 531, Linköping Studies in Science and Technology*, 1998.

A. Bozkurt, A. Rosen, H. Rosen, and B. Onaral. A portable near infrared spectroscopy system for bedside monitoring of newborn brain. *Biomed Eng Online*, 4(1):29, 2005.

N. Brunel. Dynamics of networks of randomly connected excitatory and inhibitory spiking neurons. *J Physiol Paris*, 94(5-6):445–63, 2000.

E. Bullmore, M. Brammer, S. C. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard, and P. Sham. Statistical methods of estimation and inference for functional mr image analysis. *Magn Reson Med*, 35(2):261–77, 1996.

R. B. Buxton, E. C. Wong, and L. R. Frank. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med*, 39(6):855–64, 1998.

R. B. Buxton, K. Uludag, D. J. Dubowitz, and T. T. Liu. Modeling the hemodynamic response to brain activation. *Neuroimage*, 23:220–233, 2004.

V. D. Calhoun, T. Adali, L. K. Hansen, J. Larsen, and J. J. Pekar. ICA of functional MRI data: an overview. *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, pages 281–288, 2003.

V. D. Calhoun, J. Liu, and T. Adalı. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and erp data. *Neuroimage*, 45(S1):S163–S172, 2009.

J.-F. Cardoso and A. Souloumiac. Blind beamforming for non gaussian signals. *IEE Proceedings-F*, 140:362–370, 1993.

R. Caton. The electric currents of the brain. *British Medical Journal*, 2(765):278, 1875.

Y. Chen, P. Namburi, L. T. Elliott, J. Heinzle, C. S. Soon, M. W. Chee, and J.-D. Haynes. Cortical surface-based searchlight decoding. *NeuroImage*, 56(2):582 – 592, 2011. ISSN 1053-8119. Multivariate Decoding and Brain Reading.

D. Cohen. Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents. *Science*, 161(843):784–6, 1968.

D. M. Cole, S. M. Smith, and C. F. Beckmann. Advances and pitfalls in the analysis and interpretation of resting-state fMRI data. *Front Syst Neurosci*, 4:8, 2010.

N. M. Correa, T. Eichele, T. Adalı, Y.-O. Li, and V. D. Calhoun. Multi-set canonical correlation analysis for the fusion of concurrent single trial erp and functional MRI. *Neuroimage*, 50(4):1438–1445, 2010.

T. Cover and J. Thomas. *Elements of information theory*. Wiley series in telecommunications and signal processing, 2006.

D. Cox and R. Savoy. Functional magnetic resonance imaging (fMRI)"brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, 19(2): 261–270, 2003.

B. N. Cuffin and D. Cohen. Comparison of the magnetoencephalogram and electroencephalogram. *Electroencephalogr Clin Neurophysiol*, 47(2):132–46, 1979.

F. L. da Silva and E. Niedermayer. Biophysical aspects of EEG and magnetoencephalogram generation. *Williams & Wilkins*, 1998.

A. M. Dale and E. Halgren. Spatiotemporal mapping of brain activity by integration of multiple imaging modalities. *Curr Opin Neurobiol*, 11(2):202–8, 2001.

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. *Journal of Cognitive Neuroscience*, 5: 162–176, 1993.

A. M. Dale, A. K. Liu, B. R. Fischl, R. L. Buckner, J. W. Belliveau, J. D. Lewine, and E. Halgren. Dynamic statistical parametric mapping: combining fMRI and meg for high-resolution imaging of cortical activity. *Neuron*, 26(1):55–67, 2000.

J. Daunizeau, C. Grova, G. Marrelec, J. Mattout, S. Jbabdi, M. Pélégrini-Issac, J.-M. Lina, and H. Benali. Symmetrical event-related EEG/fMRI information fusion in a variational bayesian framework. *Neuroimage*, 36(1):69–87, 2007.

H. P. O. de Beeck. Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? *Neuroimage*, 49(3):1943–8, 2010.

J. C. de Munck, S. I. Gonçalves, T. J. C. Faes, J. P. A. Kuijer, P. J. W. Pouwels, R. M. Heethaar, and F. H. L. da Silva. A study of the brain's resting state based on alpha band power, heart rate and fMRI. *Neuroimage*, 42(1):112–21, 2008.

S. Debener, M. Ullsperger, M. Siegel, and A. K. Engel. Single-trial EEG-fMRI reveals the dynamics of cognitive function. *Trends Cogn Sci*, 10(12):558–63, 2006.

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):353–9, 2003.

A. Devor, I. Ulbert, A. K. Dunn, S. N. Narayanan, S. R. Jones, M. L. Andermann, D. A. Boas, and A. M. Dale. Coupling of the cortical hemodynamic response to cortical and thalamic neuronal activity. *Proc Natl Acad Sci USA*, 102(10):3822–7, 2005.

K. I. Diamantaras and S. Y. Kung. *Principal component neural networks: theory and applications*. Wiley, 1996.

R. Duda, P. Hart, and D. Stork. Pattern classification. *Wiley*, 2001.

A. K. Dunn, A. Devor, A. M. Dale, and D. A. Boas. Spatial extent of oxygen metabolism and hemodynamic changes during functional activation of the rat somatosensory cortex. *Neuroimage*, 27(2): 279–90, 2005.

T. Eichele, K. Specht, M. Moosmann, M. L. A. 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 USA*, 102(49):17798–803, 2005.

T. Eichele, V. D. Calhoun, M. Moosmann, K. Specht, M. L. A. Jongsma, R. Q. Quiroga, H. Nordby, and K. Hugdahl. Unmixing concurrent EEG-fMRI with parallel independent component analysis. *Int J Psychophysiol*, 67(3):222–34, 2008.

S. Fazli, J. Mehnert, G. Curio, A. Villringer, K.-R. Müller, J. Steinbrink, and B. Blankertz. Enhanced performance by a hybrid NIRS-EEG brain computer interface. *Neuroimage*, 2011.

C. Fischer, F. Dailler, and D. Morlet. Novelty p3 elicited by the subject's own name in comatose patients. *Clin Neurophysiol*, 119(10):2224–30, 2008.

R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

O. Friman, M. Borga, P. Lundberg, and H. Knutsson. Exploratory fMRI analysis by autocorrelation maximization. *Neuroimage*, 16(2):454–464, 2002.

K. J. Friston. Modalities, modes, and models in functional neuroimaging. *Science*, 326(5951):399–403, 2009.

K. J. Friston and C. Buchel. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elseviers Science Publishers, 2003.

K. J. Friston, P. Jezzard, and R. Turner. Analysis of functional MRI time-series. *Human Brain Mapping*, 1:153–171, 1994.

K. J. Friston, A. P. Holmes, J. B. Poline, P. J. Grasby, S. C. Williams, R. S. Frackowiak, and R. Turner. Analysis of fMRI time-series revisited. *Neuroimage*, 2(1):45–53, 1995.

K. J. Friston, A. Mechelli, R. Turner, and C. J. Price. Nonlinear responses in fMRI: the balloon model, volterra kernels, and other hemodynamics. *Neuroimage*, 12(4):466–477, 2000.

K. J. Friston, L. Harrison, and W. Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–302, 2003.

R. D. Frostig, E. E. Lieke, D. Y. Ts'o, and A. Grinvald. Cortical functional architecture and local coupling between neuronal activity and the microcirculation revealed by in vivo high-resolution optical imaging of intrinsic signals. *Proc Natl Acad Sci USA*, 87(16):6082–6, 1990.

K. Fukumizu, F. R. Bach, and A. Gretton. Statistical consistency of kernel cca. *Journal of Machine Learning Research*, 8:361–383, 2007.

C. Fyfe. Two methods for sparsifying probabilistic canonical correlation analysis. *Proceedings of the International Conference on Neural Information Processing*, pages 1–10, 2006.

C. Fyfe and P. Lai. ICA using kernel canonical correlation analysis. *In Proc. Int. Workshop on Independent Component Analysis*, 2000.

C. Fyfe and G. Leen. Stochastic processes for canonical correlation analysis. *Proceedings of the 14th European Symposium on Artificial Neural Networks*, 2006.

P. Geladi and B. R. Kowalski. Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, 1986.

H. Girouard and C. Iadecola. Neurovascular coupling in the normal brain and in hypertension, stroke, and alzheimer disease. *Journal of Applied Physiology*, 100(1):328, 2006.

J. B. M. Goense and N. K. Logothetis. Laminar specificity in monkey v1 using high-resolution se-fMRI. *Magnetic Resonance Imaging*, 24(4):381–92, 2006.

J. B. M. Goense and N. K. Logothetis. Neurophysiology of the BOLD fMRI signal in awake monkeys. *Curr Biol*, 18(9):631–40, 2008.

R. I. Goldman, J. Stern, J. EngelJr, and M. Cohen. Acquiring simultaneous EEG and functional MRI. *Clinical Neurophysiology*, 111(11):1974–1980, 2000.

R. I. Goldman, J. M. Stern, J. Engel, and M. S. Cohen. Simultaneous EEG and fMRI of the alpha rhythm. *Neuroreport*, 13(18):2487–92, 2002.

J. Gotman, C. Grova, A. Bagshaw, E. Kobayashi, Y. Aghakhani, and F. Dubeau. Generalized epileptic discharges show thalamocortical activation and suspension of the default state of the brain. *Proc Natl Acad Sci USA*, 102(42):15236–40, 2005.

C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.

M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc Natl Acad Sci USA*, 100(1):253–8, 2003.

A. Gretton, R. Herbrich, A. J. Smola, O. Bousquet, and B. Schölkopf. Kernel methods for measuring independence. *Journal of Machine Learning Research*, 6:2075–2129, 2005.

A. Grinvald, E. Lieke, R. D. Frostig, C. D. Gilbert, and T. N. Wiesel. Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature*, 324(6095):361–4, 1986.

A. Grinvald, E. E. Lieke, R. D. Frostig, and R. Hildesheim. Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *J Neurosci*, 14(5 Pt 1):2545–68, 1994.

F. Grynszpan and D. B. Geselowitz. Model studies of the magnetocardiogram. *Biophys J*, 13(9):911–25, 1973.

R. Guibert, C. Fonta, and F. Plouraboué. Cerebral blood flow modeling in primate cortex. *J Cereb Blood Flow Metab*, 30(11):1860–73, 2010.

E. Hamel. Perivascular nerves and the regulation of cerebrovascular tone. *Journal of applied physiology*, 100(3):1059, 2006.

D. A. Handwerker, J. M. Ollinger, and M. D'Esposito. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, 21(4): 1639–51, 2004.

D. R. Hardoon and J. Shawe-Taylor. Sparse canonical correlation analysis. *Machine Learning*, 2011.

D. R. Hardoon, J. Mourão-Miranda, M. Brammer, and J. Shawe-Taylor. Unsupervised analysis of fMRI data using kernel canonical correlation. *Neuroimage*, 37(4):1250–9, 2007.

J. V. 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(5539):2425, 2001.

J.-D. Haynes and G. Rees. Predicting the stream of consciousness from activity in human visual cortex. *Current Biology*, 15(14):1301–7, 2005.

J.-D. Haynes and G. Rees. Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7(7):523–34, 2006.

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(4):323–328, 2007.

R. Henson. Demystifying parametric analyses: Illustrating canonical correlation analysis as the multivariate general linear model. *Multiple Linear Regression Viewpoints*, 2000.

A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117:500–544, 1952.

M. Hollander and D. Wolfe. Nonparametric statistical methods. *John Wiley & Sons, Inc.*, 1999.

H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3):321–377, 1936.

D. Hubel and T. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, 148:574–591, 1959.

D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160:106–54, 1962.

A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

C. Iadecola. Neurovascular regulation in the normal brain and in alzheimer's disease. *Nature Reviews Neuroscience*, 5(5):347–360, 2004.

C.-H. Im, A. Gururajan, N. Zhang, W. Chen, and B. He. Spatial resolution of EEG cortical source imaging revealed by localization of retinotopic organization in human primary visual cortex. *Journal of Neuroscience Methods*, 161(1):142–54, 2007.

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):417–20, 1993.

K. Jann, R. Wiest, M. Hauf, K. Meyer, C. Boesch, J. Mathis, G. Schroth, T. Dierks, and T. Koenig. Bold correlates of continuously fluctuating epileptic activity isolated by independent component analysis. *Neuroimage*, 42(2):635–48, 2008.

F. F. Jöbsis. Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science*, 198(4323):1264–7, 1977.

C. Jordan. Essai sur la Géometrie à *n* dimensions. *Bull. Soc. Math. France*, 3:103–174, 1875. = *Oeuvres* (Tome III), Gauthiers-Villars, Paris, 1962, 79-149.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

Y. Kamitani and Y. Sawahata. Spatial smoothing hurts localization but not information: Pitfalls for brain mappers. *Neuroimage*, 2009.

K. Katanoda, Y. Matsuda, and M. Sugishita. A spatio-temporal regression model for the analysis of functional MRI data. *Neuroimage*, 17(3):1415–28, 2002.

S. Katzner, I. Nauhaus, A. Benucci, V. Bonin, D. L. Ringach, and M. Carandini. Local origin of field potentials in visual cortex. *Neuron*, 61(1):35–41, 2009.

A. L. Keller, B. Weber, and N. K. Logothetis. Systematic investigation of vascular corrosion casts of the macaque monkey brain. *Proceedings of the 6th Forum of European Neuroscienc (FENS)*, 2008.

J. R. Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451, 1971.

S. J. Kiebel and K. J. Friston. Statistical parametric mapping for event-related potentials: I. generic considerations. *Neuroimage*, 22(2):492–502, 2004.

T. Kim, S. Wong, and R. Cipolla. Tensor canonical correlation analysis for action classification. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007.

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 (Pt 9):1679–88, 1999.

N. Kriegeskorte, R. Goebel, and P. Bandettini. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, 103(10):3863, 2006.

N. Kriegeskorte, W. K. Simmons, P. S. F. Bellgowan, and C. I. Baker. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci*, 12(5):535–540, 2009.

J. T. Kwok and I. W. Tsang. The pre-image problem in kernel methods. *Proceedings of the Twentieth International Conference on Machine Learning*, 2003.

N. Lange, S. C. Strother, J. R. Anderson, F. A. Nielsen, A. P. Holmes, T. Kolenda, R. Savoy, and L. K. Hansen. Plurality and resemblance in fMRI data analysis. *Neuroimage*, 10(3 Pt 1):282–303, 1999.

W. Larimore. Canonical variate analysis in identification, filtering, and adaptive control. *Proceedings of the 29th Conference on Decision and Control*, pages 596–604, 1990.

H. Laufs. Endogenous brain oscillations and related networks detected by surface EEG-combined fMRI. *Human Brain Mapping*, 29(7):762–769, 2008.

H. Laufs and J. S. Duncan. Electroencephalography/functional MRI in human epilepsy: what it currently can and cannot do. *Curr Opin Neurol*, 20(4):417–23, 2007.

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 USA*, 100(19):11053–8, 2003.

H. Laufs, J. Daunizeau, D. W. Carmichael, and A. Kleinschmidt. Recent advances in recording electrophysiological data simultaneously with magnetic resonance imaging. *Neuroimage*, 40(2):515–528, 2008.

A.-M. Legendre. *Nouvelles méthodes pour la détermination des orbites des comètes*, chapter Sur la methode des moindres quarres. Firmin Didot, http://imgbase-scd-ulp.u-strasbg.fr/displayimage.php?pos=-141297, 1805.

G. Lehev and M. A. L. Nicolelis. State-of-the-art microwire array design for chronic neural recordings in behaving animals–methods for neural ensemble recordings. *in Nicolelis MAL, (Ed) Methods for Neural Ensemble Recordings. 2nd edition. Boca Raton (FL): CRC Press Frontiers in Neuroscience.*, 2008.

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):943–52, 1997.

L. Lemieux, A. Salek-Haddadi, O. Josephs, P. J. Allen, N. Toms, C. Scott, K. Krakow, R. Turner, and D. Fish. Event-related fMRI with simultaneous and continuous EEG: Description of the method and initial case report. *Neuroimage*, 14(3):780–787, 2001.

S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller. Introduction to machine learning for brain imaging. *Neuroimage*, 56(2):387–99, 2011.
in press.

D. A. Leopold, Y. Murayama, and N. K. Logothetis. Very slow activity fluctuations in monkey visual cortex: Implications for functional brain imaging. *Cerebral Cortex*, pages 1–12, 2003.

J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Knowledge Discovery in Databases*, 2009.

S. Leurgans, R. Moyeed, and B. Silverman. Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society*, 1993.

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 USA*, 95(15):8945–50, 1998.

Z. Liu and B. He. fMRI-EEG integrated cortical source imaging by use of time-variant spatial constraints. *Neuroimage*, 39(3):1198–214, 2008.

N. K. Logothetis. The underpinnings of the BOLD functional magnetic resonance imaging signal. *J Neurosci*, 23(10):3963–71, 2003.

N. K. Logothetis. What we can do and what we cannot do with fMRI. *Nature*, 453(7197):869–878, 2008.

N. K. Logothetis and B. A. Wandell. Interpreting the BOLD signal. *Annu Rev Physiol*, 66:735–769, 2004.

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

N. K. Logothetis, Y. Murayama, M. A. Augath, T. Steffen, J. Werner, and A. Oeltermann. How not to study spontaneous activity. *Neuroimage*, 45(4):1080–1089, 2009.

N. K. Logothetis, M. A. Augath, Y. Murayama, A. Rauch, F. Sultan, J. B. M. Goense, A. Oeltermann, and H. Merkle. The effects of electrical microstimulation on cortical signal propagation. *Nature Neuroscience*, 13(10):1283–91, 2010.

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.

J. H. Macke, G. Zeck, and M. Bethge. Receptive fields without spike-triggering. In *Advances in Neural Information Processing Systems 20: Proceedings of the 2007 Conference*, pages 969–976, 2008.

D. Mantini, M. G. Perrucci, C. D. Gratta, G. L. Romani, and M. Corbetta. Electrophysiological signatures of resting state networks in the human brain. *Proc Natl Acad Sci USA*, 104(32):13170–5, 2007.

J. Martindale, J. Mayhew, J. Berwick, M. Jones, C. Martin, D. Johnston, P. Redgrave, and Y. Zheng. The hemodynamic impulse response to a single neural event. *J Cereb Blood Flow Metab*, 23(5):546–55, 2003.

E. Martínez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen. Concurrent EEG/fMRI analysis by multiway partial least squares. *Neuroimage*, 22(3):1023–34, 2004.

F. D. Martino, A. W. de Borst, G. Valente, R. Goebel, and E. Formisano. Predicting EEG single trial responses with simultaneous fMRI and relevance vector machine regression. *Neuroimage*, 56(2): 826–36, 2011.

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

T. Melzer, M. Reiter, and H. Bischof. Nonlinear feature extraction using generalized canonical correlation analysis. *International Conference for Artificial Neural Networks*, pages 353–360, 2001.

S. Mika, B. B. Schölkopf, A. Smola, K.-R. Müller, M. Scholz, and G. Rätsch. Kernel pca and de–noising in feature spaces. In *Advances in Neural Information Processing Systems*, pages 536–542, 1999.

M. M. Monti. Statistical analysis of fMRI time-series: A critical review of the glm approach. *Front Hum Neurosci*, 5:28, 2011.

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):145–58, 2003.

M. Moosmann, T. Eichele, H. Nordby, K. Hugdahl, and V. D. Calhoun. Joint independent component analysis for simultaneous EEG-fMRI: principle and simulation. *Int J Psychophysiol*, 67(3):212–21, 2008.

V. B. Mountcastle. Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of Neurophysiology*, 20(4):408–34, 1957.

J. Mourão-Miranda, K. J. Friston, and M. Brammer. Dynamic discrimination analysis: a spatial-temporal svm. *Neuroimage*, 36(1):88–99, 2007.

T. Muehlemann, D. Haensse, and M. Wolf. Wireless miniaturized in-vivo near infrared imaging. *Opt Express*, 16(14):10323–30, 2008.

C. Mulert and L. Lemieux. Eeg-fMRI: Physiological basis, technique, and applications. *Springer*, 2010.

K.-R. Müller, S. Mika, G. Ratsch, K. Tsuda, and B. B. Schölkopf. An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12(2):181–201, 2001.

K.-R. Müller, C. Anderson, and B. Birch. Linear and nonlinear methods for brain computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(2):165–169, 2003.

S. Murakami and Y. Okada. Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. *Journal of Physiology*, 575(3):925-36, 2006.

Y. Murayama, F. Bießmann, F. C. Meinecke, K.-R. Müller, M. A. Augath, A. Oeltermann, and N. K. Logothetis. Relationship between neural and hemodynamic signals during spontaneous activity studied with temporal kernel cca. *Magnetic Resonance Imaging*, 28(8):1095–1103, 2010.

J. Niessing, B. Ebisch, K. E. Schmidt, M. Niessing, W. Singer, and R. A. W. Galuske. Hemodynamic signals correlate tightly with synchronized gamma oscillations. *Science*, 309(5736):948–951, 2005.

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 Sciences*, 10(9):424–30, 2006.

P. L. Nunez and R. Srinivasan. Electric fields of the brain: the neurophysics of EEG. *Oxford University Press*, 2006.

A. Oeltermann, M. A. Augath, and N. K. Logothetis. Simultaneous recording of neuronal signals and functional nmr imaging. *Magnetic Resonance Imaging*, 25(6):760–74, 2007.

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

S. Ogawa, D. W. Tank, R. Menon, J. M. Ellermann, S. G. Kim, H. Merkle, and K. Ugurbil. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci USA*, 89(13):5951–5, 1992.

T. F. Oostendorp and A. van Oosterom. Source parameter estimation in inhomogeneous volume conductors of arbitrary shape. *IEEE Trans Biomed Eng*, 36(3):382–91, 1989.

D. Ostwald, C. Porcaro, and A. P. Bagshaw. An information theoretic approach to EEG-fMRI integration of visually evoked responses. *Neuroimage*, 49(1):498–516, 2010.

D. Ostwald, C. Porcaro, and A. P. Bagshaw. Voxel-wise information theoretic EEG-fMRI feature integration. *Neuroimage*, pages 1–17, 2011.

K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2: 559–572, 1901.

G. Pfurtscheller, B. Z. Allison, C. Brunner, G. Bauernfeind, T. Solis-Escalante, R. Scherer, T. O. Zander, G. Mueller-Putz, C. Neuper and N. Birbaumer The hybrid BCI. *Front Neurosci*, 4:-34, 1901.

S. M. Plis, V. D. Calhoun, M. P. Weisend, T. Eichele, and T. Lane. Meg and fMRI fusion for non-linear estimation of neural and BOLD signal changes. *Frontiers in neuroinformatics*, 4:114, 2010.

N. Pouratian, A. F. Cannestra, N. A. Martin, and A. W. Toga. Intraoperative optical intrinsic signal imaging: a clinical tool for functional brain mapping. *Neurosurg Focus*, 13(4):e1, 2002.

Karl Popper. Logik der Forschung, 1934.

M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman. A default mode of brain function. *Proc Natl Acad Sci USA*, 98(2):676–82, 2001.

C. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2005.

A. Rauch, X. Zhang, F. Bießmann, F. C. Meinecke, J. B. M. Goense, K.-R. Müller, G. Rainer, and N. K. Logothetis. Baseline BOLD signal shift in macaque primary visual cortex (V1) after local application of acetylcholine. *in preparation*, 2009.

J. J. Riera, E. Aubert, K. Iwata, R. Kawashima, X. Wan, and T. Ozaki. Fusing EEG and fMRI based on a bottom-up model: inferring activation and effective connectivity in neural masses. *Philos Trans R Soc Lond, B, Biol Sci*, 360(1457):1025–41, 2005.

J. J. Riera, J. C. Jimenez, X. Wan, R. Kawashima, and T. Ozaki. Nonlinear local electrovascular coupling. ii: From data to neuronal masses. *Human Brain Mapping*, 28(4):335–54, 2007.

P. Ritter and A. Villringer. Simultaneous EEG-fMRI. *Neurosci Biobehav Rev*, 30(6):823–38, 2006.

P. Ritter, M. Moosmann, and A. Villringer. Rolandic alpha and beta EEG rhythms' strengths are inversely related to fMRI-BOLD signal in primary somatosensory and motor cortex. *Hum Brain Mapp*, 30(4):1168–1187, 2009.

L. R. Robinson, P. J. Micklesen, D. L. Tirschwell, and H. L. Lew. Predictive value of somatosensory evoked potentials for awakening from coma. *Crit Care Med*, 31(3):960–7, 2003.

N. Roche-Labarbe, B. Zaaimi, P. Berquin, A. Nehlig, R. Grebe, and F. Wallois. Nirs-measured oxy- and deoxyhemoglobin changes associated with EEG spike-and-wave discharges in children. *Epilepsia*, 49(11):1871–1880, 2008.

M. J. Rosa, J. Daunizeau, and K. J. Friston. Eeg-fMRI integration: A critical review of biophysical modeling and data analysis approaches. *J. Int. Neurosci.*, 09(04):453, 2010a.

M. J. Rosa, J. Kilner, F. Blankenburg, O. Josephs, and W. Penny. Estimating the transfer function from neuronal activity to BOLD using simultaneous EEG-fMRI. *Neuroimage*, 49(2):1496–1509, 2010b.

K. Rosenkranz and L. Lemieux. Present and future of simultaneous EEG-fMRI. *Magn Reson Mater Phy*, 23(5-6):309–316, 2010.

S. Roweis and Z. Ghahramani. A unifying review of linear gaussian models. *Neural Computation*, 11: 305–345, 1999.

K. S. Saleem, J. M. Pauls, M. Augath, T. Trinath, B. A. Prause, T. Hashikawa, and N. K. Logothetis. Magnetic resonance imaging of neuronal connections in the macaque monkey. *Neuron*, 34(5): 685–700, 2002.

A. Salek-Haddadi, K. Friston, L. Lemieux, and D. Fish. Studying spontaneous EEG activity with fMRI. *Brain research reviews*, 43(1):110–133, 2003.

M. Sato, T. Yoshioka, S. Kajihara, K. Toyama, N. Goda, K. Doya, and M. Kawato. Hierarchical bayesian estimation for meg inverse problem. *Neuroimage*, 23(3):806–26, 2004.

R. Scheeringa, M. C. M. Bastiaansen, K. M. Petersson, R. Oostenveld, D. G. Norris, and P. Hagoort. Frontal theta EEG activity correlates negatively with the default mode network in resting state. *Int J Psychophysiol*, 67(3):242–251, 2008.

B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2002.

B. Schölkopf, A. J. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(6):1299–1319, 1998.

B. Schölkopf, S. Mika, C. Burges, P. Knirsch, K.-R. Müller, G. Rätsch and A. Smola. Input space versus feature space in kernel-based methods. *Neural Networks, IEEE Transactions on*, 10(5):1000 – 1017, 1999.

C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423, 1948.

J. Shawe-Taylor and N. Cristianini. *Kernel methods for pattern analysis*. Cambridge University Press, 2004.

A. Shmuel, E. Yacoub, D. Chaimow, N. K. Logothetis, and K. Ugurbil. Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 tesla. *Neuroimage*, 35(2):539–52, 2007.

Y. B. Sirotin, E. M. C. Hillman, C. Bordier, and A. Das. Spatiotemporal precision and hemodynamic mechanism of optical point spreads in alert primates. *Proc Natl Acad Sci U S A*, 106(43): 18390–18395, 2009.

S. M. Smith, P. T. Fox, K. L. Miller, D. C. Glahn, P. M. Fox, C. E. Mackay, N. Filippini, K. E. Watkins, R. Toro, A. R. Laird, and C. F. Beckmann. Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci USA*, 106(31):13040–5, 2009.

F. Steinke and B. Schölkopf. Kernels, regularization and differential equations. *Pattern Recognition*, 41:3271–3286, 2008.

G. Strang. Introduction to linear algebra. *Wellesley-Cambridge Press*, 2009.

S. Szedmak, T. D. Bie, and D. R. Hardoon. A metamorphosis of canonical correlation analysis into multivariate maximum margin learning. *Proceedings of the 15th European Symposium on Artificial Neural Networks*, pages 1–6, 2007.

R. Thornton, H. Laufs, R. Rodionov, S. Cannadathu, D. Carmichael, S. Vulliemoz, A. Salek-Haddadi, A. McEvoy, S. Smith, and S. Lhatoo. Eeg correlated functional MRI and postoperative outcome in focal epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, 81(8):922, 2010a.

R. C. Thornton, R. Rodionov, H. Laufs, S. Vulliemoz, A. Vaudano, D. Carmichael, S. Cannadathu, M. Guye, A. McEvoy, S. Lhatoo, F. Bartolomei, P. Chauvel, B. Diehl, F. D. Martino, R. D. C. Elwes, M. C. Walker, J. S. Duncan, and L. Lemieux. Imaging haemodynamic changes related to seizures: comparison of EEG-based general linear model, independent component analysis of fMRI and intracranial EEG. *Neuroimage*, 53(1):196–205, 2010b.

P. Tian, I. C. Teng, L. D. May, R. Kurz, K. Lu, M. Scadeng, E. M. C. Hillman, A. J. D. Crespigny, H. E. D'Arceuil, J. B. Mandeville, J. J. A. Marota, B. R. Rosen, T. T. Liu, D. A. Boas, R. B. Buxton, A. M. Dale, and A. Devor. Cortical depth-specific microvascular dilation underlies laminar differences in blood oxygenation level-dependent functional MRI signal. *Proc Natl Acad Sci USA*, 107(34): 15246–51, 2010.

M. E. Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.

M. Ullsperger and S. Debener. Simultaneous EEG and fMRI: Recording, analysis, and application. *Oxford Univ Press*, 2011.

P. A. Valdes-Sosa, J. M. Sanchez-Bornot, R. C. Sotero, Y. Iturria-Medina, Y. Aleman-Gomez, J. Bosch-Bayard, F. Carbonell, and T. Ozaki. Model driven EEG/fMRI fusion of brain oscillations. *Human Brain Mapping*, 30(9):2701–21, 2009.

V. G. van de Ven, E. Formisano, D. Prvulovic, C. H. Roeder, and D. E. J. Linden. Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. *Human Brain Mapping*, 22(3):165–78, 2004.

M. A. J. van Gerven, B. Cseke, F. P. de Lange, and T. Heskes. Efficient bayesian multivariate fMRI analysis using a sparsifying spatio-temporal prior. *Neuroimage*, 50(1):150–61, 2010.

A. Villringer and B. Chance. Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences*, 20(10):435–42, 1997.

U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

E. Vul, C. Harris, P. Winkielman, and H. Pashler. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, 4(3):274–290, 2008.

S. Vulliemoz, R. Thornton, R. Rodionov, D. W. Carmichael, M. Guye, S. Lhatoo, A. W. McEvoy, L. Spinelli, C. M. Michel, J. S. Duncan, and L. Lemieux. The spatio-temporal mapping of epileptic networks: combination of EEG-fMRI and EEG source imaging. *Neuroimage*, 46(3):834–43, 2009.

S. Vulliemoz, L. Lemieux, J. Daunizeau, C. M. Michel, and J. S. Duncan. The combination of EEG source imaging and EEG-correlated functional MRI to map epileptic networks. *Epilepsia*, 51(4): 491–505, 2010a.

S. Vulliemoz, R. Rodionov, D. W. Carmichael, R. Thornton, M. Guye, S. D. Lhatoo, C. M. Michel, J. S. Duncan, and L. Lemieux. Continuous EEG source imaging enhances analysis of EEG-fMRI in focal epilepsy. *Neuroimage*, 49(4):3219–29, 2010b.

S. Vulliemoz, D. W. Carmichael, K. Rosenkranz, B. Diehl, R. Rodionov, M. C. Walker, A. W. McEvoy, and L. Lemieux. Simultaneous intracranial EEG and fMRI of interictal epileptic discharges in humans. *Neuroimage*, 54(1):182–90, 2011.

S. Waldert, T. Pistohl, C. Braun, T. Ball, A. Aertsen, and C. Mehring. A review on directional information in neural signals for brain-machine interfaces. *Journal of Physiology Paris*, 103(3-5):244–254, 2009.

B. Weber, A. L. Keller, J. Reichold, and N. K. Logothetis. The microvascular system of the striate and extrastriate visual cortex of the macaque. *Cereb Cortex*, 18(10):2318–30, 2008.

K. Whittingstall and N. K. Logothetis. Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. *Neuron*, 64(2):281–289, 2009.

A. Wiesel, M. Kliger, and A. Hero. A greedy approach to sparse canonical correlation analysis. *http://arxiv.org/abs/0801.2748v1*, 2008.

H. Wold. *Encyclopedia of statistical sciences*, volume 6, chapter Partial least squares, pages 581–591. Wiley, 1985.

M. Wolf, M. Ferrari, and V. Quaresima. Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications. *J Biomed Opt*, 12(6):062104, 2007.

M. T. Wong-Riley. Cytochrome oxidase: an endogenous metabolic marker for neuronal activity. *Trends in Neurosciences*, 12(3):94–101, 1989.

M. W. Woolrich, M. Jenkinson, J. M. Brady, and S. M. Smith. Fully bayesian spatio-temporal modeling of fMRI data. *IEEE transactions on medical imaging*, 23(2):213–31, 2004.

K. J. Worsley and K. J. Friston. Analysis of fMRI time-series revisited–again. *Neuroimage*, 2(3):173–81, 1995.

G. Wu, X. Duan, W. Liao, Q. Gao, and H. Chen. Kernel canonical-correlation granger causality for multiple time series. *Phys Rev E Stat Nonlin Soft Matter Phys*, 83(4 Pt 1):041921, 2011.

L. Wu, T. Eichele, and V. D. Calhoun. Reactivity of hemodynamic responses and functional connectivity to different states of alpha synchrony: A concurrent EEG-fMRI study. *Neuroimage*, 52(4): 1252–1260, 2010.

D. Xing, C.-I. Yeh, and R. M. Shapley. Spatial spread of the local field potential and its laminar variation in visual cortex. *Journal of Neuroscience*, 29(37):11540–11549, 2009.

E. Yacoub, K. Ugurbil, and N. Harel. The spatial dependence of the poststimulus undershoot as revealed by high-resolution BOLD- and CBV-weighted fMRI. *J Cereb Blood Flow Metab*, 26(5): 634–44, 2006.

M. Yamada and M. R. Azimi-Sadjadi. Nonlinear signal estimation using kernel wiener filter in canonical correlation analysis framework. In *Proceedings of International Conference on Computational Intelligence for Modelling, Control and Automation*, 2005.

L. Yang, Z. Liu, and B. He. Eeg-fMRI reciprocal functional neuroimaging. *Clin Neurophysiol*, 121(8): 1240–50, 2010.

M. Ystad, E. Hodneland, S. Adolfsdottir, J. Haász, A. J. Lundervold, T. Eichele, and A. Lundervold. Cortico-striatal connectivity and cognition in normal aging: a combined dti and resting state fMRI study. *Neuroimage*, 55(1):24–31, 2011.

M. Zhao, M. Suh, H. Ma, C. Perry, A. Geneslaw, and T. H. Schwartz. Focal increases in perfusion and decreases in hemoglobin oxygenation precede seizure onset in spontaneous human epilepsy. *Epilepsia*, 48(11):2059–67, 2007.