Decomposition Methods for the Fusion of Multimodal Neuroimaging Data

Sven Dähne

Decomposition Methods for the Fusion of Multimodal Neuroimaging Data

vorgelegt von
M.Sc.
Sven Dähne
geb in Belzig

von der Fakultät IV – Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
genehmigte Dissertation

Prüfungsausschuss

Vorsitzender:	Prof. Dr. Benjamin Blankertz
1. Gutachter:	Prof. Dr. Klaus-Robert Müller
2. Gutachter:	Prof. Dr. Arno Villringer
3. Gutachter:	Dr. Vadim Nikulin

Tag der wissenschaftlichen Aussprache: 08.07.2015

Berlin 2015

Abstract

Multimodal neuroimaging, which is the study of brain function by combining (or *fusing*) several measurement modalities, holds great promise for clinical as well as cognitive neuroscience. Measurements of oscillatory brain activity are of particular interest for this thesis, because changes in this kind of neural activity (also referred to as band-power modulations) have been linked to practically every aspect of cognition and play an important role in clinical settings. However, current state-of-the-art methods for the extraction and multimodal fusion of such oscillatory neural activity fall short in a number of aspects which so far have been largely overlooked by the neuroimaging community.

This thesis contributes to multimodal fusion in a number of ways. Firstly, we present a thorough investigation of the effectiveness of popular analysis approaches by means of theoretical considerations as well as numerical simulations. These investigations are conducted with particular focus on the accuracy of the analysis approaches as well as the interpretability of their results. Only methods excelling in both of these aspects can be expected to be maximally useful for advancing our understanding of brain function. Based on these considerations, a novel multivariate analysis approach for the extraction of amplitude modulated oscillatory sources is derived. This approach is called *Source Power Co-modulation* (SPoC) analysis and it is specifically designed to overcome the shortcomings of existing methods. The utility of SPoC is thoroughly tested using numerical simulations as well as real-world neuroimaging data.

Based on the ideas that led to the SPoC analysis, we then derive two inherently multimodal extensions that allow for a principled approach to the fusion of source amplitude dynamics. The first of the two extensions is referred to as *multimodal Source Power Co-modulation* (mSPoC). mSPoC is shown to outperform existing methods in numerical simulations as well as on real-world data in the task of fusing electroencephalography (EEG) with hemodynamic measurements obtained from functional near-infrared spectroscopy (fNIRS) as well as functional magnetic resonance imaging (fMRI). The second extension of the SPoC approach is called *canonical Source Power Co-modulation* (cSPoC) and is designed for the fusion of two (or more) sets of oscillatory data. The utility of cSPoC is demonstrated in numerical simulations as well as using real-world data examples involving the fusion of (i) datasets from several subjects, (ii) different oscillatory processes within subjects, and (iii) simultaneously measured EEG and magnetoencephalography (MEG).

All of the methods developed in this thesis are based on a widely accepted linear generative model for neuroimaging data. By adhering to this model, the methods are able to uncover physiologically plausible generators of neural activity in a principled manner, thereby providing the neuroimaging community with an attractive alternative to many existing 'ad-hoc' approaches.

Zusammenfassung

Multimodale bildgebende Verfahren kombinieren (oder auch *fusionieren*) verschiedene Messmethoden zur Analyse von Hirnaktivität und beinhalten dadurch ein immenses Potential für die kognitiven und klinischen Neurowissenschaften. Für diese Dissertation sind insbesondere Fusionsszenarien von Interesse die oszillatorische Hirnaktivität mit Amplitudenmodulationen beinhalten. Gerade diese Aktivität spielt eine wichtige Rolle in klinischen Untersuchungen und steht darüber hinaus mit fast allen Aspekten der Kognition in Beziehung. Trotz der zentralen Rolle dieser Hirnprozesse, weisen viele gängige Analysemethoden Schwächen auf, welche bisher weithin übersehen wurden.

An dieser Stelle setzt die vorliegende Dissertation an und trägt durch folgende Beiträge zum Fortschritt des Forschungsfeldes bei. Zunächst werden die gängigen Analysemethoden sowohl durch analytische Überlegungen als auch durch numerische Simulationen gründlich auf zwei wichtige Aspekte hin untersucht. Diese Aspekte sind die Genauigkeit und die Interpretierbarkeit der von den Methoden gelieferten Ergebnisse. Nur Methoden die mit Hinblick auf beide Aspekten überzeugen, können von maximalen Nutzen für die Erforschung des Gehirns sein. Basierend auf den Ergebnissen dieser Untersuchung wird eine neue Methode vorgestellt, welche die vorher identifizierten Schwächen überwindet. Ihr Name ist *Source Power Co-modulation* (SPoC) Analyse.

Basierend auf den Überlegungen die zur SPoC Analyse geführt haben, werden zwei multimodale Erweiterungen vorgestellt: die *multimodal Source Power Co-modulation* (mSPoC) Analyse und die *canonical Source Power Co-modulation* (cSPoC) Analyse. Konkrete Anwendungsbeispiele die für mSPoC vorgestellt und diskutiert werden sind die Fusion von (i) Elektroenzophalografie (EEG) und funktioneller Nah-infrarot Spektroskopie (fNIRS) sowie (ii) EEG und funktioneller Magnetresonanztomografie (fMRT). cSPoC wurde speziell für die Fusion von mehreren amplitudenmodulierten oszillatorischen Datensätzen entwickelt. Die Nützlichkeit von cSPoC wird anhand von insgesamt drei Analyse-/Fusionsszenarien gezeigt: (i) die Fusion der Daten von mehreren Versuchspersonen, (ii) die Fusion verschiedener oszillatorischer Prozesse innerhalb von Versuchspersonen und (iii) die Fusion von simultan gemessenen EEG und Magnetenzephalographie (MEG) Daten.

Alle in dieser Dissertation vorgestellten Analysemethoden basieren auf einem weitläufig akzeptierten linearen generativen Modell für das EEG und MEG. Der starke Bezug auf dieses Modell bildet die theoretisch fundierte Grundlage der Methoden. Die praktische Relevanz der Methoden spiegelt sich wider in der erfolgreichen Extraktion von physiologisch plausiblen Hirnaktivitätsmustern aus einer Vielzahl von Beispieldaten. Dadurch stellen die entwickelten Analysemethoden eine attraktive Alternative zu vielen oft eher “ad-hoc” wirkenden multimodalen Fusionansätzen dar.

Acknowledgements

The work presented in this thesis would not have been possible without the help and support from a number of people whom I am extremely grateful for.

Firstly, I would like to thank my supervisor Prof. Klaus-Robert Müller for providing such a wonderful work environment and for allowing me to freely explore my analysis ideas. It is his gentile guidance and valuable input that brings the right people together and creates innovation.

Next, I would like to thank Prof. Arno Villringer, Dr. Vadim Nikulin, and Prof. Benjamin Blankertz for reviewing and marking of my thesis. The collaboration with Arno's lab has been very productive and continues to lead to interesting insights. The work with Vadim has been wonderfully efficient and I hope the future holds many more successful collaborations between us. A special thanks goes to Benjamin for being so inspiring and yet so down-to-earth at the same time.

The first two years of my PhD experience were centered around the TOBI project. I am grateful for the guidance that Michael Tangermann has provided during that time. Together with Johannes Höhne and Martijn Schreuder, I worked a lot on BCI's during that time, which has taught me a great deal in terms of signal processing and machine learning. In addition to the science, we were an awesome team and had fun times together. Thank you Johannes for the uncountably many discussions about research directions, analysis methods, competitions, business ideas, and surf spots around the world. Your sometimes critical but always constructive feedback is what has kept – and continues to keep – me on my toes.

I am also very grateful for the help and advice I have gotten from what used to be the "older generation" of PhD students and Post-Docs in our lab. Stefan, Frank, Felix, Paul, Siamac, Matthias, Mikio, Shin, and Andreas – I have learned a lot from you guys and I have very much enjoyed working with you.

Thank you Bastian, Matthias, Irene W., Irene S., Laura, Markus, Daniel, Han-Jeong, Janne, and many others for being such a great team. Thank you Andrea Gerdes, Imke Weitkamp, and Dominik Kühne for your support with administrative issues and IT problems.

Finally I would like to thank the most important people in my personal life: my parents and Jee Hei. Thank you mum and dad for always believing in me and the constant support on the path that has led me to this day. Thank you Jee Hei for simply being the most awesome partner anyone could ever ask for. Top shelf!

Contents

Chapter 1

Introduction

1.1 The fusion of multimodal neuroimaging data

Modern neuroscience benefits greatly from a multitude of imaging techniques that, individually, have helped to further our understanding of cognitive processing (Eichele et al., 2005; Debener et al., 2006) and improved clinical diagnostics (Dale and Halgren, 2001; Vulliemoz et al., 2011). The combination of several imaging modalities originated in the context of epilepsy imaging (Ives et al., 1993; Lemieux et al., 1997; Daunizeau et al., 2007) but has since then become an important asset in cognitive neuroscience. It is only through multimodal setups that otherwise unparalleled spatial and temporal imaging resolution can be obtained. This, in turn, allows for combination of complementary information and thereby for a better diagnosis and a deeper understanding of how different aspects of brain activity are related.

The most popular multimodal imaging setups combine measurements of electrophysiology with measurements of hemodynamics. Example techniques for measuring electrophysiological properties of neural activity are electrocorticography (ECoG), electroencephalography (EEG), and magnetoencephalography (MEG). Examples for techniques that measure changes in hemodynamic parameters include positron emission tomography (PET), functional near-infrared spectroscopy (fNIRS), and functional magnetic resonance imaging (fMRI). See reviews on electrophysiological aspects of brain activity (e.g. Niedermeyer and da Silva, 2005; Buzsáki et al., 2012) or hemodynamic aspects aspects of brain activity (e.g. Attwell and Iadecola, 2002; Arthurs and Boniface, 2002) as well as chapter 2.1 for more details on neuroimaging measurement devices and the types of signals they record.

The task of optimal combination of information from several (imaging) modalities is referred to as *multimodal analysis* or *multimodal fusion*. Multimodal fusion includes a number of exquisite challenges for signal processing and machine learning. Therefore, multimodal fusion still represents an ongoing research endeavor, as there is no gold standard solution (Bießmann et al., 2011; Huster et al., 2012; Dähne et al., 2015).

The following list enumerates some of the key challenges that make multimodal fusion a difficult problem:

- **Different spatial and temporal sampling rates:** The number of recording channels typically range from approximately one hundred for electrophysiology to near one million voxels for hemodynamics. The picture is reversed, however, for temporal

sampling rates where electrophysiology is typically sampled in the kHz range while hemodynamics are sampled with rates below 10 Hz.

- **Non-instantaneous and non-linear coupling:** The vascular reaction to a given stimulus is in the range of seconds, while the response in electrophysiological measures (e.g. event-related potentials (ERPs)) occur in the range of milliseconds. Furthermore, non-linear features such as the instantaneous amplitude of neural oscillations may be related to linear features of hemodynamics.
- **The presence of outliers and low signal-to-noise ratio (SNR):** Signals of interest may not be easily detectable at the level of individual measurement channels due to a low SNR. The existence of outliers in the data, (either caused by technological or physiological artifacts) may further shadow the signals of interest and lead estimates of certain statistics of the data astray and thereby hinder successful fusion.
- **Interpretable results:** The aim of multimodal imaging settings is to increase our understanding of the workings of the brain. Therefore, the results of multimodal fusion techniques must be interpretable with respect to functional or anatomical neurophysiological references.

In order to overcome these challenges it is helpful to regard multimodal fusion as modeling as well as an optimization problem. With respect to both of these two views one class of statistical learning methods has become particularly popular for multimodal data analysis: *factor models*.

These models assume that the measurements are the result of the activity of a limited set of *components* (see chapter 2.2 for the formal definition) of which a mixture is observed at the level of the sensors of the measurement device. Un-mixing these components, requires a set of assumptions about the nature of the components. Different assumptions lead to different statistical learning methods and therefore it is important to know these assumptions when choosing an analysis method.

1.2 How this thesis contributes to multimodal fusion

In this thesis, we present a set of novel methods that each extract a set of components from multivariate uni- as well as multimodal measurements. These methods are distinguished from earlier approaches by two key properties. Firstly the methods presented here assume a functional relation (or *functional coupling*) between the modalities. This functional coupling is assumed to be reflected in a temporal co-modulation of activity. Secondly, the methods are specifically designed for the extraction of oscillatory signals whose *amplitude modulations* (or *bandpower modulations*) exhibit the temporal co-modulation with another modality and thereby mediate the functional coupling.

Chapter 2 covers the fundamentals on which this thesis is based. The chapter begins by reviewing basic facts about the two main sources of neuroimaging data, namely electrophysiological and hemodynamic measurements. Thereafter, in the same chapter we review a linear generative model for neuroimaging recordings and introduce the concepts of source *components* and how to extract their *spatial activation patterns* as well as temporal profiles

(or *time-courses*) from the data using *spatial filters*. The chapter concludes by reviewing a set of state-of-the-art decomposition methods for multivariate data. These methods represent the baseline to which the novel approaches presented in the following chapters will be compared.

The novel contributions of this thesis begin with **chapter 3**, which provides a detailed account of some of the shortcomings that current state-of-art methods suffer from. These shortcomings are then addressed in **chapter 4**, which introduces the *Source Power Co-modulation analyis* (SPoC), the first of three novel approaches derived in this thesis. Thereafter, in **chapter 5**, the *multimodal Source Power Co-modulation analyis* (mSPoC) is presented, which specifically addresses the fusion of amplitude modulations in electrophysiological recordings with hemodynamic signals. The fusion of amplitude modulations from different electrophysiological signal sources (i.e. separate measurement modalities or multiple individuals) is addressed in **chapter 6**, in which the *canonical Source Power Co-modulation* (cSPoC) analysis is presented. Finally, this thesis concludes with a more general discussion in **chapter 7**.

Included publications

The work presented in this thesis has been published in peer-reviewed journals and conferences. In fact, in this thesis we follow those publications very closely. The following lists enumerates the publications that constitute the main body of this thesis. The publications are listed in chronological order, starting with the most recent.

1. Dähne, S., Bießmann, F., Samek, W., Haufe, S., Goltz, D., Gundlach, C., Villringer, A., Fazli, S., and Müller, K.-R. (2015). Multivariate machine learning methods for fusing functional multimodal neuroimaging data. *Proceedings of the IEEE*, 103(9):1507 – 1530

2. Dähne, S., Nikulin, V. V., Ramírez, D., Schreier, P. J., Müller, K.-R., and Haufe, S. (2014e). Finding brain oscillations with power dependencies in neuroimaging data. *NeuroImage*, 96:334–348

3. Dähne, S., Meinecke, F. C., Haufe, S., Höhne, J., Tangermann, M., Müller, K.-R., and Nikulin, V. V. (2014d). SPoC: a novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. *NeuroImage*, 86(0):111–122

4. Dähne, S., Bießman, F., Meinecke, F. C., Mehnert, J., Fazli, S., and Müller, K.-R. (2013). Integration of multivariate data streams with bandpower signals. *IEEE Transactions on Multimedia*, 15(5):1001–1013

All publications

This subsection lists all peer-reviewed publications that I have (co-)authored between 2010 and 2015. The items are ordered chronologically, starting with the most recent. Journal articles that are not yet published are given in a separate list. The current status of each item in that list is given as a note at the end of the entry.

Journal articles (published or accepted for publication)

1. Sturm, I., Treder, M., Miklody, D., Purwins, H., Dähne, S., Blankertz, B., and Curio, G. (2015b). The polyphonic brain: Extracting the neural representation of tone onsets for separate voices of polyphonic music using multivariate EEG analysis. *Psychomusicology*. accepted

2. Venthur, B., Dähne, S., Höhne, J., Heller, H., and Blankertz, B. (2015). Wyrm: A brain-computer interface toolbox in Python. *Journal of Neuroinformatics*, pages 1–16

3. Dähne, S., Bießmann, F., Samek, W., Haufe, S., Goltz, D., Gundlach, C., Villringer, A., Fazli, S., and Müller, K.-R. (2015). Multivariate machine learning methods for fusing functional multimodal neuroimaging data. *Proceedings of the IEEE*, 103(9):1507 – 1530

4. Fazli, S., Dähne, S., Samek, W., Bießmann, F., and Müller, K.-R. (2015). Learning from more than one data source: data fusion techniques for sensorimotor rhythm-based Brain-Computer Interfaces. *Proceedings of the IEEE*, 103(6):891–906

5. Hahne, J., Dähne, S., Hwang, H.-J., , Müller, K.-R., and Parra, L. C. (2015). Concurrent adaptation of human and machine improves simultaneous and proportional myoelectric control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. accepted for publication

6. Winkler, I., Haufe, S., Porbadnigk, A., Müller, K.-R., and Dähne, S. (2015). Identifying Granger causal relationships between neural power dynamics and variables of interest. *NeuroImage*, 111:489–504

7. Haufe, S., Dähne, S., and Nikulin, V. V. (2014a). Dimensionality reduction for the analysis of brain oscillations. *NeuroImage*, 101:583–597

8. Dähne, S., Nikulin, V. V., Ramírez, D., Schreier, P. J., Müller, K.-R., and Haufe, S. (2014e). Finding brain oscillations with power dependencies in neuroimaging data. *NeuroImage*, 96:334–348

9. Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., and Bießmann, F. (2014b). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, 87:96–110. Neuroimage single best paper of 2014 Award

10. Dähne, S., Wilbert, N., and Wiskott, L. (2014g). Slow feature analysis of retinal waves leads to V1 complex cells. *PLoS Computational Biology*, 10(5):e1003564

11. Dähne, S., Meinecke, F. C., Haufe, S., Höhne, J., Tangermann, M., Müller, K.-R., and Nikulin, V. V. (2014d). SPoC: a novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. *NeuroImage*, 86(0):111–122

12. Dähne, S., Bießman, F., Meinecke, F. C., Mehnert, J., Fazli, S., and Müller, K.-R. (2013). Integration of multivariate data streams with bandpower signals. *IEEE Transactions on Multimedia*, 15(5):1001–1013

13. Martel, A., Dähne, S., and Blankertz, B. (2014). EEG predictors of covert vigilant attention. *Journal of Neural Engineering*, 11(3):035009

14. Schreuder, M., Riccio, A., Risetti, M., Dähne, S., Ramsey, A., Williamson, J., Mattia, D., and Tangermann, M. (2013). User-centered design in BCI - a case study. *Artificial Intelligence in Medicine*, 59(2):71–80

15. Höhne, J., Krenzlin, K., Dähne, S., and Tangermann, M. (2012). Natural stimuli improve auditory BCIs with respect to ergonomics and performance. *Journal of Neural Engineering*, 9(4):045003

16. Venthur, B., Scholler, S., Williamson, J., Dähne, S., Treder, M. S., Kramarek, M. T., Müller, K.-R., and Blankertz, B. (2010). Pyff – a pythonic framework for feedback applications and stimulus presentation in neuroscience. *Frontiers in Neuroscience*, 4:179

17. Weiller, D., Märtin, R., Dähne, S., Engel, A. K., and König, P. (2010). Involving motor capabilities in the formation of sensory space representations. *PLOS ONE*, 5(4):e10377

Journal articles (not yet published because currently under review or in preparation)

18. Schultze-Kraft*, M., Dähne*, S., Blankertz, B., and Curio, G. (2015). Supervised and unsupervised classification of visuomotor workload states with EEG. in preparation

19. Sturm, I., Dähne, S., Blankertz, B., and Curio, G. (2015a). Multi-variate EEG analysis as a novel tool to examine brain responses to naturalistic music stimuli. *PLOS ONE*. in revision

20. Wong, D.-O., Hwang, H.-J., Dähne, S., Müller, K.-R., and Lee, S.-W. (2015). Effect of higher frequency on steady state visual evoked potential based brain-computer interface. *Journal of Neural Engineering*. in revision

Conference contributions and abstracts (selection)

1. Hansen, S. T., Winkler, I., Hansen, L. K., Müller, K.-R., and Dähne, S. (2015). Fusing simultaneous EEG and fMRI using functional and anatomical information. In *International Workshop on Pattern Recognition in Neuroimaging, 2015*. IEEE. accepted

2. Dähne, S., Haufe, S., Bießmann, F., Meinecke, F., Ramirez, D., Schreier, P., Nikulin, V., and Müller, K.-R. (2014c). Finding brain oscillations with power dependencies in neuroimaging data. In *Annual Meeting of the Organization for Human Brain Mapping (OHBM), 2014*

3. Dähne, S., Nikulin, V. V., Ramirez, D., Schreier, P. J., Müller, K.-R., and Haufe, S. (2014f). Optimizing spatial filters for the extraction of envelope-coupled neural oscillations. In *International Workshop on Pattern Recognition in Neuroimaging, 2014*. IEEE

4. Dähne, S., Hahne, J., Pawletta, P., and Müller, K.-R. (2014b). Boosting simultaneous and proportional myoelectric control by combining source power correlation (spoc) and linear regression. In *Bernstein Conference, 2014*

5. Dähne, S., Bießmann, F., Meinecke, F. C., Mehnert, J., Fazli, S., and Müller, K.-R. (2014a). Multimodal integration of electrophysiological and hemodynamic signals. In *Brain-Computer Interface (BCI), 2014 International Winter Workshop on*, pages 1–4

6. Horn, F. and Dähne, S. (2013). Increasing the spectral signal-to-noise ratio of common spatial patterns. In *5th International BCI Meeting in Asilomar*

7. Dähne, S., Meinecke, F., Haufe, S., Höhne, J., Tangermann, M., Nikulin, V., and Müller, K.-R. (2012). Multi-variate correlation of power spectral density. In *Annual Meeting of the Organization for Human Brain Mapping (OHBM), 2012*

8. Dähne, S., Höhne, J., Schreuder, M., and Tangermann, M. (2011a). Band power features correlate with performance in auditory brain-computer interface. In *Front. Hum. Neurosci. Conference Abstract: XI International Conference on Cognitive Neuroscience (ICON XI)*

9. Dähne, S., Höhne, J., Schreuder, M., and Tangermann, M. (2011b). Slow feature analysis - a tool for extraction of discriminating event-related potentials in brain-computer interfaces. In Honkela, T., Duch, W., Girolami, M., and Kaski, S., editors, *Artificial Neural Networks and Machine Learning - ICANN 2011*, volume 6791 of *Lecture Notes in Computer Science*, pages 36–43. Springer Berlin / Heidelberg

10. Dähne, S., Höhne, J., and Tangermann, M. (2011c). Adaptive classification improves control performance in ERP-based BCIs. In *Proceedings of the 5th International BCI Conference*, pages 92–95, Graz

Chapter 2

Fundamentals

2.1 Measures of brain activity

The brain is one of the most complicated – yet at the same time most fascinating – natural systems currently under scientific investigation. And although considerable progress has been made in understanding many of the basic principles that govern brain activity, there is still a multitude of aspects and phenomena that are far from being fully understood.

The functional nature of the brain can be quantified (and thereby analyzed) using a number of different measurement approaches. The two approaches to neuroimaging that are most relevant for this thesis are *electrophysiological-* and *hemodynamic* measures, which we now briefly review.

2.1.1 Electrophysiological measures

Neural activity results in changes of electrical fields (Buzsáki et al., 2012), which can be measured at various spatial, temporal and functional extents (Scanziani and Häusser, 2009). Intracellular recordings allow for measuring action potentials of single neurons (Hodgkin and Huxley, 1939). The activity of single and multiple neurons up to large neuronal assemblies can be extracted with extracellular recording techniques, either invasively with microelectrodes inserted in the brain or ECoG (Wyler et al., 1984) or non-invasively with EEG (Berger, 1929) or MEG (Cohen, 1968).

Extracellularly measured local field potentials (LFP) represent a superposition of all currents in the brain, with a distance-dependent contribution of different sources such as synaptic currents, calcium-spikes, action potentials and spike afterpotentials of different neurons (Buzsáki et al., 2012). While signals measured from microelectrodes and ECoGs can represent rather focal and localized signatures of neuronal activity, signals measured with EEG rely on synchronous activity of large assemblies of neurons. Such synchronous activity is often resembled in neuronal oscillations (Buzsáki and Draguhn, 2004) and the spatial synchronization strength is reflected in the power of these oscillations (Denker et al., 2011; Pfurtscheller and Lopes da Silva, 1999). Neural oscillations have been linked to practically every aspect of cognitive function (see below for examples) and are thus also the subject of multimodal analysis settings. Besides these neuronal oscillations, there is synchronized activity of neurons measurable with electrophysiological methods that follows certain events

or the presentation of stimuli. Such activity is often termed as event related potentials (ERPs), with different components attributed to various cognitive processes (Regan, 1989; Luck, 2014). In this thesis however, we will focus on bandpower modulations of neural oscillations.

Cognitive phenomena, that have been shown to correlate with bandpower modulations, include attention (Başar et al., 1997; Debener et al., 2003; Bauer et al., 2006), memory encoding (Klimesch, 1999; Jensen et al., 2007), vigilance in operational environments (Gevins et al., 1995; Holm et al., 2009), sleep stages (Darchia et al., 2007; Demanuele et al., 2012), perception (Plourde et al., 1991; Makeig and Jung, 1996; Thut et al., 2006), and decision making (Haegens et al., 2011a,b), for example. In Transcranial Magnetic Stimulation research it has also been shown that the excitability of the motor (Sauseng et al., 2009) and visual cortices (Romei et al., 2008), measured in terms of the amplitude of alpha oscillations, can also be predictive of muscular motor evoked potentials and visual perception, respectively. In the field of Brain-Computer Interfaces (BCI), voluntary modulation of EEG band power is used to control computer applications, such as text entry systems (Blankertz et al., 2007, 2008). It has been shown recently that variability in BCI control-performance can be partially explained by the variability of spectral power across subjects (Blankertz et al., 2010; Suk et al., 2014), as well as within subjects (Grosse-Wentrup et al., 2011; Dähne et al., 2011a; Maeder et al., 2012).

All of the aforementioned studies investigated the relation between (the power of) neural oscillations and certain external variables or experimental conditions. However, it is noteworthy that interactions between oscillatory processes have also been of increasing interest in the recent past (Jensen and Colgin, 2007; Canolty and Knight, 2010; Hipp et al., 2012; Engel et al., 2013). Different types of so-called *cross-frequency coupling* between neural systems have been identified and associated with specific brain functions. Such couplings can be divided into phase-to-frequency coupling, phase-to-phase coupling, phase-to-power coupling, and finally power-to-power coupling (see Jensen and Colgin, 2007, for an overview). Power-to-power (or amplitude-to-amplitude) coupling, which will be of special interest later in this thesis, describes the interaction between the spectral power (or the envelope/amplitude) of distinct oscillations at specific frequencies or narrow frequency bands.

2.1.2 Hemodynamic measures

Hemodynamic activity can be measured invasively by intrinsic optical imaging (Grinvald et al., 1986). Non-invasive alternatives exist in the form of fMRI (Kwong et al., 1992; Ogawa et al., 1990) or fNIRS (Jobsis, 1977).

fMRI measures the combination of metabolic and vascular response to neural activation, the so-called blood oxygen-level dependent (BOLD) signal (Ogawa et al., 1990). The BOLD signal is inversely related to the local concentration of deoxygenated hemoglobin (HbR), which in turn is influenced by changes in cerebral blood volume (CBV) and cerebral blood flow (CBF) (Buxton et al., 2004). Since HbR is paramagnetic, while oxygenated hemoglobin (HbO) is not, only changes in the concentration of HbR alter the local magnetic susceptibility

and hence give rise to the fMRI signal obtained in a magnetic-resonance (MR) scanner with so-called T2*-weighted pulse sequences.

Functional near infrared spectroscopy (fNIRS) relies on the fact that near-infrared light can traverse biological tissue and thus allows the transmission of photons through the intact brain (Jobsis, 1977; Thorniley et al., 1990). The absorption properties of HbR and HbO differ substantially in the infrared range (Wray et al., 1988). This enables to measure changes in concentrations of HbR and HbO *in vivo*. When compared to fMRI, fNIRS measurements can be performed with a lightweight and comparatively low-cost setup. Similar to EEG, light emitting and detection devices (so-called *optodes*) can be mounted on a fNIRS cap.

The relationship between neural activity and the vascular response is known as *neurovascular coupling* (Attwell and Iadecola, 2002; Buxton et al., 2004; Hamel, 2006; Logothetis and Wandell, 2004) and the exact nature of this coupling is far from understood (Heeger and Ress, 2002; Logothetis, 2008; Hall et al., 2014). However, recently a number of studies have shown that neural and hemodynamic signals are highly correlated (Berwick et al., 2008; Bonvento et al., 2002; Devor et al., 2005; Goense and Logothetis, 2008; Logothetis et al., 2001; Logothetis and Wandell, 2004; Martindale et al., 2003; Sirotin et al., 2009). Simultaneous intracranial electrophysiological recordings and high-resolution fMRI in macaque monkeys, for example, revealed a correlation between the BOLD signal and neuronal activity in the gamma range as a neurovascular coupling mechanism (Goense and Logothetis, 2008; Logothetis et al., 2001). Similar results have been obtained in cats (Niessing et al., 2005). However, neurovascular coupling can also be assessed using noninvasive methods such as combined EEG-fNIRS (Ritter et al., 2009; Fazli et al., 2012) or EEG-fMRI (Moosmann et al., 2003; Laufs et al., 2003; Mantini et al., 2007; Laufs et al., 2008). These and other studies have demonstrated an inverse relationship between the BOLD signal and the amplitude of neural oscillations in the alpha and beta range as well as a peak in correlation at a time delay of 6 to 8 seconds.

2.2 A multimodal generative model

We now review an abstract mathematical model for the generation of macroscopic neuroimaging data, such as EEG, MEG, NIRS and fMRI. The essence of this model is that the measured data can be represented as linear mixture of functionally distinct processes, where each of theses processes itself is characterized by a specific spatial as well as temporal signature. These processes are referred to as *components* and the generative model expresses the measurements in terms of the spatial and temporal patterns of the underlying components.

From here on we represent the neuroimaging datasets to be fused by the symbols $\mathbf{x}$ and $\mathbf{y}$. A single observation is denoted by $\mathbf{x}(t) \in \mathbb{R}^{N_\mathbf{x}}$ and $\mathbf{y}(t) \in \mathbb{R}^{N_\mathbf{y}}$, where $N_\mathbf{x}$ and $N_\mathbf{y}$ denote the number of recording channels in each modality. The matrices that contain all data points are denoted by $\mathbf{X} = [\mathbf{x}(1), ..., \mathbf{x}(T_\mathbf{x})] \in \mathbb{R}^{N_\mathbf{x} \times T_\mathbf{x}}$ and $\mathbf{Y} = [\mathbf{y}(1), ..., \mathbf{y}(T_\mathbf{y})] \in \mathbb{R}^{N_\mathbf{y} \times T_\mathbf{y}}$. Further symbols used in this thesis and their meaning is summarized in Table 2.1.

Note that in general $N_\mathbf{x} \neq N_\mathbf{y}$, i.e. the two datasets are not required to have the same number of channels. In fact, in practical settings it is most likely that the measurement

modalities have been sampled drastically different in space as well as in time. However, we do assume that the datasets represent the same time line, meaning that the same events have been observed in the same chronological order. In practice, it is likely for the datasets to have been recorded simultaneously during the course of an experiment and thus the "same time line" assumption is fulfilled trivially. However, this assumption can also be fulfilled by datasets that have been recorded consecutively, rather than simultaneously, using the exact same stimulus sequence, for example.

$T_{\mathbf{x/y}}$	Number of data points, modality specific
$N_{\mathbf{x/y}}$	Number of measurement channels, modality specific
$K_{\mathbf{x/y}}$	Number of components (i.e. latent factors), modality specific
K	Number of coupled components, i.e. components that are linked across modalities
$\mathbf{x}(t)$	$N_{\mathbf{x}}$-dimensional vector of observed data in modality $\mathbf{x}$
$\mathbf{y}(t)$	$N_{\mathbf{y}}$-dimensional vector of observed data in modality $\mathbf{y}$
$\mathbf{X}, \mathbf{Y}$	$N_{\mathbf{x/y}} \times T_{\mathbf{x/y}}$ matrix containing the observed data, modality specific
$\mathbf{s}_{\mathbf{x/y}}(t), \hat{\mathbf{s}}_{\mathbf{x/y}}(t)$	$K_{\mathbf{x/y}}$-dimensional column-vector of (estimated) components
$\boldsymbol{\epsilon}_{\mathbf{x/y}}(t)$	$N_{\mathbf{x/y}}$-dimensional noise vector in forward models
$\mathbf{A}_{\mathbf{x/y}}$	$N_{\mathbf{x/y}} \times K_{\mathbf{x/y}}$ matrix of sensor-space activation patterns in forward models
$\mathbf{W}_{\mathbf{x/y}}$	$N_{\mathbf{x/y}} \times K_{\mathbf{x/y}}$ matrix of filters in backward models
$\mathbf{C}_{\mathbf{xx/yy}}$	Data covariance, modality specific
$\mathbf{C}_{\mathbf{xy/xy}}$	Cross-modal data covariance matrix

Table 2.1: Notation.

2.2.1 Forward model

The central assumption we make is that a given dataset is decomposable into what is called a set of *components* (or *factors*). The notion of a component underlies all of the state-of-the-art as well as the newly developed models presented in this thesis. Thus, we define what a component is in more detail in the following paragraphs.

An individual component is identified by a unique temporal and spatial signature and may thus be regarded as a functional unit. The component notation is congruent for $\mathbf{x}$ and $\mathbf{y}$, so we introduce the notation exemplary for $\mathbf{x}$ only. Let the scalar variable $s^i_{\mathbf{x}}(t)$ denote the temporal signature of a component with the index i at time point t. We will also refer to $s^i_{\mathbf{x}}(t)$ as the *temporal activity* of this component. The strength with which $s^i_{\mathbf{x}}(t)$ is expressed

at each recording channel is called the *spatial activation pattern* and denoted by the vector[1] $\mathbf{a}^i_{\mathbf{x}} \in \mathbb{R}^{N_{\mathbf{x}}}$.

Generally a given dataset is assumed to be composed of a set of $K_{\mathbf{x}} \geq 1$ components. Let $\mathbf{s}_{\mathbf{x}}(t) \in \mathbb{R}^{K_{\mathbf{x}}}$ denote the vector which represents the temporal activity of the $K_{\mathbf{x}}$ components and let $\mathbf{A}_{\mathbf{x}} \in \mathbb{R}^{N_{\mathbf{x}} \times K_{\mathbf{x}}}$ denote the matrix in which each column contains the corresponding spatial activation patterns. We will make use of these variables when we consider the mapping from components (i.e. hidden variables) to recordings (i.e. observable variables). This mapping is referred to as the *linear forward model* or *linear generative model*. In this model, the projection of the components to the recording channels is given by

$$\begin{aligned} \mathbf{x}(t) &= \sum_{i=1}^{K_{\mathbf{x}}} \mathbf{a}^i_{\mathbf{x}} \cdot s^i_{\mathbf{x}}(t) + \boldsymbol{\epsilon}_{\mathbf{x}}(t) \\ &= \mathbf{A}_{\mathbf{x}}\mathbf{s}_{\mathbf{x}}(t) + \boldsymbol{\epsilon}_{\mathbf{x}}(t) \,, \end{aligned} \tag{2.1}$$

where $\boldsymbol{\epsilon}_{\mathbf{x}}(t) \in \mathbb{R}^{N_{\mathbf{x}}}$ captures activity that is not explained by the $K_{\mathbf{x}}$ components and thus considered *noise*. The task of machine learning is to extract estimates of the underlying components from the data. We use $\hat{\mathbf{s}}_{\mathbf{x}}$ and $\hat{\mathbf{s}}_{\mathbf{y}}$ to denote these estimates. The above equation describes the observed data as a linear superposition of components, each of which is characterized by a time course and a fixed spatial activation pattern.

The datasets $\mathbf{x}$ and $\mathbf{y}$ are assumed to be related by $K \leq \min(K_{\mathbf{x}}, K_{\mathbf{y}})$ pairs of shared component processes among the rows of $\hat{\mathbf{s}}_{\mathbf{x}}$ and $\hat{\mathbf{s}}_{\mathbf{y}}$. The exact nature of the relation between the shared components of course depends on the measurement modalities that are being used. However, a very generic connection between the modalities can be constructed based on the assumption that the datasets to represent a common timeline and thereby provide different views upon the same underlying processes. Therefore, it is to be expected that the time courses of shared components (or certain features of these time courses) exhibit "similar" dynamics. We formalize this notion of similarity by the following expression:

$$\Phi(\hat{s}^i_{\mathbf{x}}) \approx_{\mathrm{r}} \Psi(\hat{s}^i_{\mathbf{y}}), \tag{2.2}$$

for $i \in \{1, ..., K\}$. The functions $\Phi(\cdot)$ and $\Psi(\cdot)$ extract some feature from the time course of the component pair $\hat{s}^i_{\mathbf{x}}$ and $\hat{s}^i_{\mathbf{y}}$ that is similar in terms of a similarity metric $\approx_{\mathrm{r}}$. Examples for the feature extracting functions $\Phi(\cdot)$ and $\Psi(\cdot)$ could be simply the identity function, a function extracting spectral features, (de-)convolution operators, or functions extracting statistical properties of the distributions of $\hat{s}^i_{\mathbf{x}}$ or $\hat{s}^i_{\mathbf{y}}$. Examples for similarity measuring functions are functions that measure co-modulation in time, such as *covariance* or *correlation*. Another popular choice for the similarity metric is an information theoretic measures that is called *mutual information*. Note that in contrast to covariance and correlation, mutual information captures nonlinear dependencies between variables.

Fusion approaches A decomposition into components can be done separately for each modality or jointly for all measurement modalities. We refer to the former approach as *late fusion* scenarios and to the latter approach as *early fusion*. In this thesis we focus on

[1] Please note that i is not the exponent of the variable but denotes the ith component.

methods for these two approaches to the analysis of multimodal functional neuroimaging data.

While in principle all of the presented models can be extended to more than two modalities, we here focus – for the sake of readability – on the special case of two different measurement modalities. Furthermore, since the scope of this thesis is limited to *functional* neuroimaging data, we assume all measurements to be temporally aligned.

Figure 2.1 summarizes the notions presented in this subsection: it outlines the generative model of multimodal neuroimaging data that is adopted here and it contrasts early and late approaches to multimodal fusion.

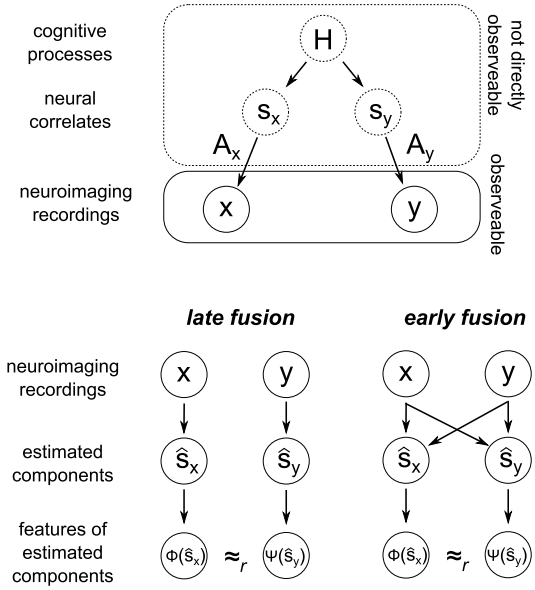


Figure 2.1: A multimodal generative model (top) and two generic fusion approaches to multimodal data (bottom). A cognitive phenomenon (**H**, e.g. attention, stimulus processing) influences certain aspects of modality specific neurophysiological processes, such as electrophysiological or metabolic properties. In the context of this generative model, these processes are modeled by latent variables (also called sources) and denoted by $\mathbf{s_{x/y}}$. These latent variables are mapped by a modality specific transformation ($\mathbf{A_{x/y}}$) to their respective sensor space variables ($\mathbf{X}/\mathbf{Y}$). Starting from the recorded datasets $\mathbf{X}$ and $\mathbf{Y}$, it is the task of factor model based methods to extract estimates of the latent sources ($\hat{\mathbf{s}}_{\mathbf{x/y}}$) such that features of the estimated source activity ($\Phi(\hat{\mathbf{s}}_{\mathbf{x}})$ and $\Psi(\hat{\mathbf{s}}_{\mathbf{y}})$) are informative about $\mathbf{H}$ itself, or tell us something about how exactly $\mathbf{H}$ exerts influence on $\mathbf{s_{x/y}}$. In *early fusion* approaches, information from both modalities is already taken into account when extracting source activity from the data. In *late fusion* approaches, modality-specific sources are extracted without using information from the respective other modality first, and features of the estimated sources are combined thereafter.

2.2.2 Backward model

After having expressed the recorded data as a sum of underlying components, where each component is the product of a specific spatial and temporal signature, the question arises how to recover the components from the data. In the most general setting, the factors $\mathbf{A_x}$ and $\mathbf{s_x}$ in Eq. (2.1) are estimated jointly, a setting that is referred to as blind source separation (BSS). However, the factorization into $\mathbf{A_x}$ and $\mathbf{s_x}$ is not unique and therefore further assumptions have to be made about the nature of the spatial and temporal dynamics. As we show later, different assumptions lead to different factorization methods.

Estimating both the spatial activation patterns and the time-courses jointly leads to potentially difficult optimization problems. The computational complexity, however, can be reduced by resorting to a so-called linear discriminative (or *backward*) modeling approach. In such an approach, the time-courses of K neural sources are estimated by projecting the data linearly onto a set of spatial *extraction filters* $\mathbf{W_x} \in \mathbb{R}^{N_x \times K}$:

$$\hat{\mathbf{s}}_\mathbf{x}(t) = \mathbf{W}_\mathbf{x}^\top \mathbf{x}(t). \tag{2.3}$$

The coefficients of $\mathbf{W_x}$ determine how to integrate the information from all recording channels in order to optimally extract the time-courses of the components. Several approaches to find, or rather to optimize, these coefficients will be presented in the following chapter. However, at this point it is important to discuss some common misconceptions about the interpretability of the coefficients of filters, once they have been obtained.

2.2.3 Recovering the forward model from a backward model: patterns

Earlier we have identified "neurophysiological interpretation" as a key property desired in (multimodal) neuroimaging. A prerequisite for determining the anatomical origin and neurophysiological relevance of extracted time-courses, is to identify the strength with which the time-courses are expressed at each recording channel. Importantly however, the coefficients of extraction filters do *not* encode this information and should therefore *not* be interpreted with respect to the origin of the extracted signal. Interpretation is only possible for the activation patterns of forward models (Haufe et al., 2014b; Blankertz et al., 2011). Moreover, it is only the activation patterns that can be subjected to source localization techniques in order to link cognitive functions to specific brain areas. The reason for this is that extraction filters are generally functions of the signal *and* the noise and thus heavily influenced by factors not of interest for the neurophysiological interpretation.

See Fig. 2.2 for an illustration of the duality between filters and patterns. In the figure, the top panel on the right side shows the time-courses of two hidden source components, s_1 and s_2, of which s_1 shall be the signal of interest in this example and s_2 corresponds to a noise component. These time-courses are mapped to two recording channels (x_1 and x_2) by means of Eq. (2.1), using the matrix $\mathbf{A_x} = \left(\mathbf{a}^1, \mathbf{a}^2\right)$. The time-courses of the data in channel space is shown in the middle panel on the right side, as well as in a scatter plot on the left side (x_1 on the abscissa and x_2 on the ordinate). The scatter plot also shows the activation patterns (columns of the matrix $\mathbf{A_x}$) as solid line vectors. Note that $\mathbf{a}^1$ is only half as long as $\mathbf{a}^2$, which means that the noise component s_2 is expressed much stronger in

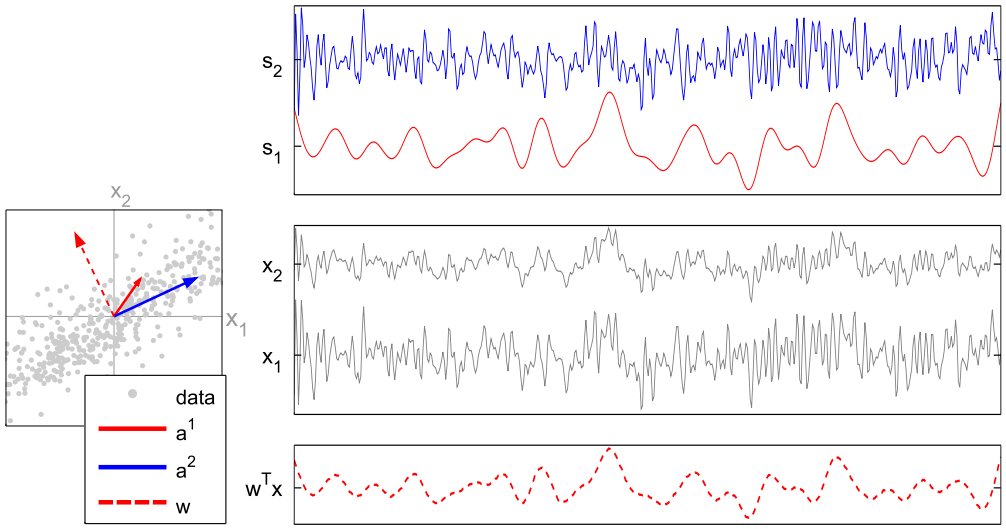


Figure 2.2: Illustration of the difference between extraction filters (i.e. the coefficients of backward models) and spatial activation patterns (i.e. the coefficients of forward models). See the main text for a detailed description.

the channel data. In this example, the weight vector that extracts s_1 best is given by the vector that is orthogonal to the activation pattern of the noise component. The optimal weight vector $\mathbf{w}$ is shown in the scatter plot as a dashed line vector. Applying $\mathbf{w}$ to the data, i.e. computing $\mathbf{w}^\top \mathbf{x}(t)$, yields a reconstruction of s_1, see the lower panel on the right side. Importantly, while $\mathbf{w}$ extracts the time-course of component s_1, its coefficients are not to be interpreted as to how strong and with what sign s_1 was expressed in the data. Instead, only the coefficients of $\mathbf{a}^1$ contain that information. However, an estimate of $\mathbf{a}^1$ can be derived from $\mathbf{w}$ by means of Eq. (2.4).

Note that in this example, $\mathbf{w}$ yields a near perfect reconstruction of $\mathbf{s}_1$ because $\mathbf{w}$ is orthogonal to $\mathbf{a}^2$. Thus the coefficients of $\mathbf{w}$ are in fact determined by the spatial pattern of the noise component $\mathbf{a}^2$, rather than by the pattern of the component of interest $\mathbf{a}^1$. This underlines the danger of interpreting filters instead of patterns.

As a remedy, a corresponding forward model of the form of Eq. (2.3) may be derived from every linear backward model (Haufe et al., 2014b). The activation patterns of the derived forward model can then be interpreted in the aforementioned way. The transformation of backward model extraction filters into forward model activation patterns is given by

$$\begin{aligned} \mathbf{A_x} &= \mathbf{C_{xx}} \mathbf{W_x} \mathbf{C}_{\hat{\mathbf{s}}_\mathbf{x} \hat{\mathbf{s}}_\mathbf{x}}^{-1} \\ &= \mathbf{C_{xx}} \mathbf{W_x} \left(\mathbf{W_x}^\top \mathbf{C_{xx}} \mathbf{W_x} \right)^{-1} , \end{aligned} \tag{2.4}$$

where $\mathbf{C_{xx}}$ denotes the data covariance matrix and $\mathbf{C}_{\hat{\mathbf{s}}_\mathbf{x} \hat{\mathbf{s}}_\mathbf{x}}$ denotes the covariance matrix of component time-courses.

By virtue of the transformation (2.4), we can pursue a backward modeling approach, allowing us to conveniently parameterize cost functions solely in terms of the extraction filters, while being able to achieve neurophysiological interpretability and source localization through analysis of the activation patterns of the corresponding forward model.

2.3 State-of-the-art analysis techniques

Next we review multivariate backward models for the extraction of components from unimodal as well as multimodal datasets. Unimodal methods are widely used in the previously introduced late fusion scenarios. In these fusion approaches, modality specific components (or features thereof) are extracted before fusion with respective features from the other modality.

Some of the state-of-the-art methods we discuss here will represent the basis of comparison against the methods that we derive in the second part of this thesis. At first we discuss unimodal methods, which we further subdivide into *supervised* and *unsupervised* methods. Supervised methods make use of an external target signal during the optimizing of the parameters, while unsupervised methods rely on the statistics of the data alone. Thereafter, we discuss inherently multimodal analysis methods.

A reoccurring theme in all the methods presented in the following is to guide the search for weight vectors $\mathbf{w}$ by means of optimizing what is called an *objective function*. For the majority of methods we will discuss below, the objective function takes the form

$$\max_{\mathbf{w}}/\min_{\mathbf{w}} \ \mathbf{w}^\top \mathbf{B}_1 \mathbf{w} \ \text{ subject to } \mathbf{w}^\top \mathbf{B}_2 \mathbf{w} = c, \tag{2.5}$$

where c is a constant and $\mathbf{B}_1$ as well as $\mathbf{B}_1$ are quadratic matrices with their number of rows and columns equal to the number of entries in $\mathbf{w}$. The aim of this generic optimization problem is to minimize (maximize) the term $\mathbf{w}^\top \mathbf{B}_1 \mathbf{w}$, subject to the constraint that $\mathbf{w}^\top \mathbf{B}_2 \mathbf{w}$ must be equal to c.

The methods we discuss below differ with respect to the specific choice of the matrices $\mathbf{B}_1$ and $\mathbf{B}_2$. However, if an objective function can be expressed in the above form, the solution is obtained as the solution to the corresponding generalized eigenvalue problem

$$\mathbf{B}_1 \mathbf{w} = \lambda \mathbf{B}_2 \mathbf{w}, \tag{2.6}$$

where λ denotes what is called the generalized *eigenvalue* that is associated with the *eigenvector* $\mathbf{w}$. Generalized eigenvalue problems have been studied for decades in the field of numerical linear algebra, which has lead to efficient algorithms for solving them (Francis, 1961; Kublanovskaya, 1962). Being able to cast an objective function into the form of a generalized eigenvalue problem is desirable because it can then be solved using standard numerical linear algebra tools such as MATLAB or R, for example.

2.3.1 Unsupervised unimodal approaches

Principal Component Analysis (PCA)

Perhaps the most popular and most widely used unsupervised factorization method is the principal component analysis (PCA, Pearson, 1901; Jolliffe, 1982). The underlying idea in PCA is to find components in the data that account for as much variance as possible under the constraint that the components are mutually de-correlated.

Let us formalize an objective for PCA for a single component. The coefficients of the weight vector $\mathbf{w}$ are to be optimized such that the extracted signal $\mathbf{w}^\top \mathbf{x}(t)$ has maximum variance:

$$\max_{\mathbf{w}} \ \mathrm{Var}\left(\mathbf{w}^\top \mathbf{x}(t)\right), \quad \text{subject to } \|\mathbf{w}\|^2 = 1 \ . \tag{2.7}$$

Expressing the variance of $\mathbf{w}^\top \mathbf{x}(t)$ as

$$\mathrm{Var}(\mathbf{w}^\top \mathbf{x}(t)) = \mathbf{w}^\top \mathbf{C} \mathbf{w}, \tag{2.8}$$

where the matrix $\mathbf{C}$ is the covariance matrix of the data, we arrive at

$$\max_{\mathbf{w}} \ \mathbf{w}^\top \mathbf{C} \mathbf{w}, \quad \text{subject to } \mathbf{w}^\top \mathbf{w} = 1. \tag{2.9}$$

which corresponds to Eq. (2.5) with $\mathbf{B}_1 = \mathbf{C}$ and $\mathbf{B}_2 = \mathbf{I}$. Thus the corresponding eigenvalue equation is given by

$$\mathbf{C}\mathbf{w} = \lambda \mathbf{w}, \tag{2.10}$$

and the solution is obtained as the eigen-decomposition of the covariance matrix $\mathbf{C}$.

In the case of PCA, the corresponding forward model takes on a particular simple form. The weight vectors $\mathbf{W}$ optimized by PCA are the eigenvectors of the covariance matrix $\mathbf{C}$, i.e. it holds that $\mathbf{C} = \mathbf{W}\Lambda\mathbf{W}^\top$, where Λ is a diagonal matrix and $\mathbf{W}\mathbf{W}^\top = \mathbf{I}$. Using the last two equations and the filter/pattern transformation presented in Eq. (2.4) we can show that in the case of PCA we have

$$\begin{aligned} \mathbf{A} &\overset{\text{Eq. (2.4)}}{=} \mathbf{C}\mathbf{W}(\mathbf{W}^\top \mathbf{C}\mathbf{W})^{-1} \\ &= \mathbf{W}\Lambda\mathbf{W}^\top \mathbf{W}(\mathbf{W}^\top \mathbf{W}\Lambda\mathbf{W}^\top \mathbf{W})^{-1} \\ &= \mathbf{W}\Lambda(\Lambda)^{-1} \\ &= \mathbf{W} \ . \end{aligned} \tag{2.11}$$

This derivation reveals that for PCA it holds that $\mathbf{A} = \mathbf{W}$, i.e. the spatial patterns of the corresponding forward model are equivalent to the spatial filters. Note however, that this is only the case for PCA, because the weight vectors that optimize the PCA objective diagonalize the covariance matrix $\mathbf{C}$, i.e. $\mathbf{C} = \mathbf{W}\Lambda\mathbf{W}^\top$.

The orthogonality of the patterns represents a very strong constraint and it is questionable that such a constrain is satisfied by physiological activation patterns. In contrast to PCA the methods discussed next do not impose the orthogonality constraint on the patterns.

Independent Component Analysis (ICA)

ICA is based on the idea that the hidden components are statistically independent. Let $\hat{\mathcal{S}}$ denote the random variable that contains the temporal signature of the extracted components. Then $\hat{\mathcal{S}}$ is parametrized by the weight matrix $\mathbf{W}$, by virtue of the backward modeling approach in Eq. (2.3). The notion of maximal independence between the individual components, denoted by the random variables $\hat{\mathcal{S}}_i$ for $i \in \{1, ..., K\}$ is equivalent to the notion of minimizing the mutual information (MI) between them.

Mathematically, the MI between two variables X and Y is defined as the Kullback-Leibler divergence D_{KL} between the joint probability distribution of X and Y (denoted as p) and the product of the marginal probability distributions

$$\begin{aligned} \mathcal{I}(X,Y) &= D_{KL}(p(X,Y) \,||\, p(X)p(Y)) \\ &= \int_X \int_Y p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy \; . \end{aligned} \tag{2.12}$$

Furthermore, the mutual information of $\hat{\mathcal{S}}_i$ can be expressed as

$$\mathcal{I}(\hat{\mathcal{S}}) = \sum_i^K \mathcal{H}(\hat{\mathcal{S}}_i) - \mathcal{H}(\hat{\mathcal{S}}), \tag{2.13}$$

where $\mathcal{H}(\hat{\mathcal{S}}_i)$ denotes the *entropy* of $\hat{\mathcal{S}}_i$. It can be shown that minimizing $\mathcal{I}(\hat{\mathcal{S}})$ can be achieved by minimizing the entropy for all individual components. Since the Gaussian distribution has the maximal entropy among distributions with fixed mean and variance, the mutual information between components can be minimized by extracting components with maximally non-Gaussian distributions. A number of algorithms exist that are based on the this idea (e.g. Bell and Sejnowski, 1995; Cardoso and Souloumiac, 1993a; Hyvarinen, 1999).

A different approach to ICA is taken by methods that exploit temporal information. These methods are based on the joint (approximate) diagonalisation of time-lagged covariance matrices. Examples are described in Belouchrani et al. (1997) and Ziehe and Müller (1998).

Note that the independence assumption used in ICA can be applied to either the estimated time courses of the components (as was outlined above) or to their estimated activation patterns. The former approach is referred to as *temporal* ICA, while the latter is called *spatial* ICA (McKeown et al., 1997). In the context of fMRI, spatial ICA is the more popular version, while in the context of EEG and MEG, temporal ICA is used. See Calhoun et al. (2009) for more discussion on the choice between spatial and temporal ICA.

Spatio-Spectral Decomposition (SSD)

Given the interest in generators of neural oscillations, we review an unsupervised backward model for the extraction of oscillatory sources next.

Modulations in the strength of a narrow-band oscillatory signal can be extracted by means of the Hilbert transform, which allows to separate the instantaneous amplitude (also called

envelope) and the instantaneous phase (Barlow, 1993). A useful approximation of the (squared) envelope is given by computing the variance of the narrow-band signal in short consecutive time windows. We refer to these time windows as *epochs* and index them with the index e.

Let $\mathbf{X}(e) \in \mathbb{R}^{N \times T_e}$ denote the matrix that contains all samples within an epoch, where the epoch is indexed by e. We assume $\mathbf{x}(t)$ to be bandpass filtered for the band of interest and thus use the variance approximation of spectral power in the given frequency band. We further denote the bandpower of $\mathbf{w}^\top \mathbf{x}(t)$ within epoch e by $\phi_{\mathbf{w}}(e)$, with

$$\phi_{\mathbf{w}}(e) = \mathrm{Var}\left(\mathbf{w}^\top \mathbf{X}(e)\right) = \mathbf{w}^\top \mathbf{C}(e)\mathbf{w}, \tag{2.14}$$

where $\mathbf{C}(e)$ denotes the covariance matrix of $\mathbf{x}$ computed for the epoch e, similar to Eq. (2.8).

It now follows that the average spectral power of a component in the band of interest is given by

$$\langle \phi_{\mathbf{w}}(e) \rangle = \left\langle \mathbf{w}^\top \mathbf{C}(e)\mathbf{w} \right\rangle = \mathbf{w}^\top \left\langle \mathbf{C}(e) \right\rangle \mathbf{w} = \mathbf{w}^\top \mathbf{C}\mathbf{w}, \tag{2.15}$$

where $\langle \cdot \rangle$ denotes the temporal average (here averaging over epochs), $\mathbf{C}(e)$ denotes the covariance matrix of $\mathbf{x}(t)$ computed for the epoch e, and $\mathbf{C}$ denotes the covariance matrix computed using all data. Note that for $\langle \mathbf{C}(e) \rangle = \mathbf{C}$ to hold, the mean of $\mathbf{x}$ within each epoch must be zero, which we assume to be the case.

The purpose of *spatio-spectral decomposition* (SSD, Nikulin et al., 2011) is to extract oscillatory signals with a high signal-to-noise ratio (SNR). The SNR is defined here as the ratio between the spectral power in the frequency band of interest and the power in the left and right neighboring (flanking) frequency bands. Using the variance approximation for power, the spectral signal-to-noise ration of the projected signal can be expressed as

$$\frac{\mathrm{Var}(\mathbf{w}^\top \mathbf{x})}{\mathrm{Var}(\mathbf{w}^\top \mathbf{x}_{\mathrm{noise}})} \quad = \quad \frac{\mathbf{w}^\top \mathbf{C}\mathbf{w}}{\mathbf{w}^\top \mathbf{C}_{\mathrm{noise}}\mathbf{w}} \,, \tag{2.16}$$

where here $\mathbf{x}$ are the measured data bandpassed filtered in the band of interest, while $\mathbf{x}_{\mathrm{noise}}$ are the data filtered in the sidebands. $\mathbf{C}$ and $\mathbf{C}_{\mathrm{noise}}$ denote the respective the covariance matrix. Practically, filtering in the left and right side bands can be implemented by applying a bandpass filter covering both side bands followed by a bandstop filter cutting out the band of interest.

Thus, the objective optimized by SSD filters is given by

$$\max_{\mathbf{w}} \; \frac{\mathbf{w}^\top \mathbf{C}\mathbf{w}}{\mathbf{w}^\top \mathbf{C}_{\mathrm{noise}}\mathbf{w}} \,. \tag{2.17}$$

Taking the derivative with respect to $\mathbf{w}$ and setting it to zero leads to

$$\lambda \mathbf{C}_{\mathrm{noise}}\mathbf{w} = \mathbf{C}\mathbf{w} \,, \tag{2.18}$$

which corresponds to the generalized eigenvalue equation we have seen in Eq. (2.5) with $\mathbf{B}_1 = \mathbf{C}$ and $\mathbf{B}_2 = \mathbf{C}_{\mathrm{noise}}$.

2.3.2 Supervised unimodal approaches

In this section we assume that in addition to the data from the imaging modality, we are also given an external target signal, denoted by the scalar variable z. This variable may encode additional information about the stimulus (e.g. type, intensity, latency, etc.), behavioral measurements (reaction times, ratings, etc.), external physiological parameters (skin conductance, heart rate, etc.), or artifactual information (e.g. eye movements, motion parameters, etc.). In general, supervised methods have an advantage over unsupervised approaches because they have more information at their disposal.

Regression and classification

Two well known examples of supervised factor models are linear regression and classification by means of linear discriminant analysis (LDA). We will first examine linear regression and then treat LDA as a special case of the former.

The goal of regression is to extract a component with a time-course that co-modulates with the target variable z. Without loss of generality, we assume z to have zero mean and unit variance. One way of quantifying co-modularity between two time series is by way of the *mean squared error* (MSE), given by

$$\mathrm{MSE}\left(\mathbf{w}^\top \mathbf{x}(t), z(t)\right) = \frac{1}{T}\sum_{t}^{T} \frac{1}{2}\left(\mathbf{w}^\top \mathbf{x}(t) - z(t)\right)^2. \tag{2.19}$$

Let the matrix $\mathbf{X} = (\mathbf{x}(1), ..., \mathbf{x}(T)) \in \mathbb{R}^{N\times T}$ contain the data and the row-vector $z = (z(1), ..., z(T)) \in \mathbb{R}^{1\times T}$ contain the time-course of the target variable. Then the spatial filter that minimizes the MSE is given by

$$\mathbf{w} = \left(\mathbf{X}\mathbf{X}^\top\right)^{-1}\mathbf{X}z^\top, \tag{2.20}$$

which is known as the ordinary least squares (OLS) solution.

Interestingly, the same solution is obtained for the following objective function, which expresses co-modularity in terms of covariance between $\mathbf{w}^\top\mathbf{x}(t)$ and $z(t)$:

$$\max_{\mathbf{w}} \ \mathrm{Cov}\left(\mathbf{w}^\top\mathbf{X}, z\right), \ \text{ subject to } \ \mathrm{Var}\left(\mathbf{w}^\top\mathbf{X}\right) = 1. \tag{2.21}$$

Or equivalently in matrix notation expressed as

$$\max_{\mathbf{w}} \ \mathbf{w}^\top\mathbf{X}z^\top, \ \text{ subject to } \ \mathbf{w}^\top\mathbf{C}\mathbf{w} = 1. \tag{2.22}$$

Linear regression is a special case of the general linear model (GLM) framework that has been successfully applied in the context of unimodal and asymmetric multimodal analysis of fMRI data for almost two decades (Friston et al., 1994). In the context of fMRI, the target variable is usually the time-course of an fMRI voxel, while $\mathbf{x}$ is called the *design matrix*. Each column of the design matrix contains a *regressor*, which are explanatory variables such

as stimulus level or task-condition for example. As this approach is usually applied to all voxels separately, it is referred to as *mass-univariate* analysis.

Classification of two conditions (or classes) can be treated within the regression framework outlined above. In such a scenario, the target variable z takes on only two values that indicate class membership. Eq. (2.20) yields the filter that achieves optimal class separation by choosing $z(t) = p(c_1)$ for all time points t that belong to class 1 and $z(t) = -p(c_2)$ for all time points t that belong to class 2, where $p(c_1)$ denotes the prior probability of class 1 and $p(c_2)$ the prior probability of class 2 (see, for example, chapter 4 in Bishop (2006)). The resulting algorithm is called linear discriminant analysis (LDA).

Common Spatial Patterns

In Brain-Computer Interfaces (BCI) based on motor imagery (MI), spatial filters are commonly applied to the recorded data in order reduce the effects of volume conduction and thereby increase the signal-to-noise ratio. This greatly aids in the detection of changes in sensorimotor activity and thus enables single-trial classification of motor-related oscillatory activity (Müller et al., 2008). A very popular and powerful spatial filtering technique in the context of MI BCIs is the so-called *common spatial patterns* (CSP) algorithm (Koles, 1991; Ramoser et al., 2000; Blankertz et al., 2008; Tomioka and Müller, 2010).

Given band-passed data from a two-class BCI paradigm, CSP optimizes spatial filters that maximize the variance of the spatially filtered signal for one class, while minimizing the variance for the respective other class. Note that, similar to what we have seen in the section on SSD, variance of the band-passed signal serves as an approximation of bandpower. Thereby it represents as a measure of event-related (de-)synchronization (ERD/ERS).

Using the variance formalism we have seen before, one can express the desired difference in variance for two conditions in the following objective function:

$$\max_{\mathbf{w}} \frac{\mathbf{w}^\top (\mathbf{C}_1 - \mathbf{C}_2)\, \mathbf{w}}{\mathbf{w}^\top (\mathbf{C}_1 + \mathbf{C}_2)\, \mathbf{w}} , \tag{2.23}$$

where $\mathbf{C}_1$ denotes the covariance matrix of all trials of class one, while $\mathbf{C}_2$ denotes the covariance matrix of class 2. The entire matrix of spatial filters can be obtained by means of the generalized eigenvalue decomposition of the matrices $(\mathbf{C}_1 - \mathbf{C}_2)$ and $(\mathbf{C}_1 + \mathbf{C}_2)$,

$$(\mathbf{C}_1 - \mathbf{C}_2)\, \mathbf{W} = (\mathbf{C}_1 + \mathbf{C}_2)\, \mathbf{W}\mathbf{D} . \tag{2.24}$$

Interestingly, the columns of $\mathbf{W}$ not only contain the spatial filter that maximizes Eq. (2.23) but also the one that minimizes it. Let the columns of $\mathbf{w}$ be ordered according to the corresponding eigenvalues in the diagonal of $\mathbf{D}$, in decreasing order. Then the first and last column of $\mathbf{w}$ are the projection vectors that maximize the difference in variances between the two classes, because the first eigenvector maximizes Eq. (2.23) and the last eigenvector minimizes it. In practice, however, not only the first and the last eigenvector are chosen but two or three from each side of the corresponding eigenvalue spectrum.

2.3.3 Multimodal methods

In this section we discuss factor models that are designed to decompose two (or more) datasets at the same time. These approaches make it possible to integrate information from both measurement modalities for the extraction of components. Since the modalities inform each other, multimodal methods have an advantage over unimodal methods. For simplicity we here assume just two modalities, denoted by $\mathbf{x}$ and $\mathbf{y}$, but the concepts presented below can be extended to more than two modalities. In the context of simultaneous measurements of electrophysiology and hemodynamics, $\mathbf{x}$ represents the former and $\mathbf{y}$ the later.

Multimodal versions of ICA

Joint ICA (jICA), presented in Calhoun et al. (2006), is a method that enables fusion of multimodal features from several of subjects. Let N_s denote the number of subjects and $\mathbf{D_x} \in \mathbb{R}^{N_s \times N_{f\mathbf{x}}}$ and $\mathbf{D_y} \in \mathbb{R}^{N_s \times N_{f\mathbf{y}}}$ denote the matrices that contain features from the $\mathbf{x}$ and $\mathbf{y}$ modality, respectively.

In the next step the features from the modalities are simply concatenated along the horizontal to yield a multimodal feature matrix $\mathbf{D} = [\mathbf{D_x}, \mathbf{D_y}] \in \mathbb{R}^{N_s \times (N_{f\mathbf{x}} + N_{f\mathbf{y}})}$. Each row in the matrix $\mathbf{D}$ corresponds to the multimodal feature concatenation of a single subject. Joint ICA now assumes the following generative model:

$$\mathbf{D} = \mathbf{G} \cdot \mathbf{V}^\top = \sum_i^K \mathbf{g}^i \cdot {\mathbf{v}^i}^\top, \tag{2.25}$$

which states that the multimodal feature matrix $\mathbf{D}$ can be decomposed into the sum of $K = \min(N_s, (N_{f\mathbf{x}} + N_{f\mathbf{y}}))$ components. Each of the components is characterized by a multimodal feature profile $\mathbf{v}^i \in \mathbb{R}^{N_{f\mathbf{x}} + N_{f\mathbf{y}}}$ and vector $\mathbf{g}^i \in \mathbb{R}^{N_s}$, for $i \in \{1, ..., N_s\}$, that encodes how strong and with which sign the feature profile is present in each of the subjects. Assuming statistical independence between the feature profiles $\mathbf{v}^i$, a backward modeling approach can be applied to extract an estimate of these profiles by ICA algorithms discussed earlier.

The natural scaling of data from different modalities, i.e. Voltage in EEG vs percent signal change or concentration changes in fMRI or fNIRS, yield quite different histograms and may thus lead methods astray that rely on information-theoretic measures. This is the case for jICA. Additionally, an unequal number of samples between the two modalities leads to jICA giving more priority to the modality for which more samples are provided. In order to ensure a balanced representation, up-/downsampling has to be applied.

While jICA assumes a common modulation profile within modalities for all subjects, this assumption is relaxed in an approach called *parallel ICA* (paraICA) (see Liu and Calhoun, 2007; Liu et al., 2009, for example). In this approach, a user specified similarity relation between components from the different modalities is optimized simultaneously with modality-specific un-mixing matrices. Thereby paraICA gives more emphasis to subject-specific multimodal components, compared to jICA.

Recently, a fully Bayesian approach to multimodal ICA was proposed in Groves et al. (2011), in which the authors presented the so-called *linked ICA*. In contrast to jICA and paraICA, a difference in scaling or noise levels between modalities is not a problem for linked ICA.

CCA

For finding related components, a useful assumption is temporal co-modulation, which can be captured by finding those transformations for each modality that maximize the correlation between the time-courses of the extracted components. This is the idea of Canonical Correlation Analysis (CCA) (Hotelling, 1936). In the simplest case CCA finds a one-dimensional subspace $\mathbf{w_x} \in \mathbb{R}^{N_\mathbf{x}}$ and $\mathbf{w_y} \in \mathbb{R}^{N_\mathbf{y}}$ for data from two modalities such that the *canonical correlation* of the modalities in that subspace is maximized:

$$\max_{\mathbf{w_x},\mathbf{w_y}} \operatorname{Corr}\left(\mathbf{w}_\mathbf{x}^\top \mathbf{x}(t), \mathbf{w}_\mathbf{y}^\top \mathbf{y}(t)\right) . \tag{2.26}$$

The advantage of maximizing the correlation after the linear transformation $\mathbf{w_x}, \mathbf{w_y}$ is that the resulting correlation coefficient is invariant with respect to linear transformations of the data, hence *canonical*. The generalization of the univariate canonical correlation coefficient finds K dimensional subspaces $\mathbf{W_x} \in \mathbb{R}^{N_\mathbf{x} \times K}$ and $\mathbf{W_y} \in \mathbb{R}^{N_\mathbf{y} \times K}$ such that the sum of the correlations is maximized (Kettenring, 1971). In matrix notation this objective can be written as

$$\begin{aligned} &\max_{\mathbf{W_x},\mathbf{W_y}} \operatorname{Trace}\left(\mathbf{W}_\mathbf{x}^\top \mathbf{X}\mathbf{Y}^\top \mathbf{W}_\mathbf{y}\right) \\ \text{subject to } \; &\mathbf{W}_\mathbf{x}^\top \mathbf{X}\mathbf{X}^\top \mathbf{W}_\mathbf{x} = \mathbf{W}_\mathbf{y}^\top \mathbf{Y}\mathbf{Y}^\top \mathbf{W}_\mathbf{y} = \mathbf{I} . \end{aligned} \tag{2.27}$$

The objective of CCA in Eq. (2.27) can be transformed into the following generalized eigenvalue problem:

$$\begin{bmatrix} 0 & \mathbf{C_{xy}} \\ \mathbf{C_{yx}} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{W_x} \\ \mathbf{W_y} \end{bmatrix} = \Lambda \begin{bmatrix} \mathbf{C_x} & 0 \\ 0 & \mathbf{C_y} \end{bmatrix} \begin{bmatrix} \mathbf{W_x} \\ \mathbf{W_y} \end{bmatrix}, \tag{2.28}$$

where $\mathbf{C_{xy}}$, $\mathbf{C_{yx}}$, $\mathbf{C_{xx}}$, $\mathbf{C_{yy}}$ are defined in Table 2.1.

If $\mathbf{C_x}$ and $\mathbf{C_y}$ are assumed to be the identity matrix, that is assuming that the features of $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated, respectively, Eq. 2.28 solves an optimization problem that is known as *partial least squares* (PLS, Sun et al., 2009; Krishnan et al., 2011), which has also found applications in multimodal data fusion (Martínez-Montes et al., 2004). The main difference between PLS and CCA is that CCA aims at finding maximally correlated components, while PLS aims at finding maximally covarying components. While this can be the same in some cases, in practice this is not necessarily so. The correct choice of method depends on what aspects of the data the analyst or experimenter wants to investigate.

CCA has been extended to handle more than two modalities at the same time. These, so-called, *multiset* extensions of CCA have found application in multimodal neuroimaging as well (Correa et al., 2008, 2009; Varoquaux et al., 2010; Gaebler et al., 2014). See also the review on CCA by Correa et al. (2010) which discusses the differences between multiset

CCA and jICA: jICA seeks to find independent components, which can be too strong an assumption in some cases. More importantly, unlike jICA, multiset CCA does not constrain the activation patterns of components to be the same for both modalities.

Temporal kernel CCA (tkCCA)

Note that CCA assumes that the samples of each modality are correlated instantaneously. For neuroimaging data this assumption does often not hold true. One solution is to embed one modality in its temporal context and optimize a time-lag-dependent projection $\mathbf{w}_\mathbf{x}(\tau)$ for one modality, such that the canonical correlation is maximized:

$$\max_{\mathbf{w}_\mathbf{x}(\tau),\mathbf{w}_\mathbf{y}} \mathrm{Corr}\left(\sum_{i}^{N_\tau}(\mathbf{w}_\mathbf{x}(\tau_i)^\top \mathbf{x}(t-\tau_i), \mathbf{w}_\mathbf{y}^\top \mathbf{y}(t)\right), \tag{2.29}$$

for a given set of N_τ time lags $\{\tau_1, ..., \tau_{N_\tau}\}$. The solution to Eq. (2.29) can be conveniently obtained as the solution to the standard CCA problem in Eq. (2.26) by applying the trick of *temporal embedding*. Temporal embedding is achieved by first creating N_τ copies of the dataset which is to be embedded (here $\mathbf{X}$), then shifting each copy by one of the specified time lags, and finally stacking the time-shifted copies along the spatial axis of the data matrix. Let the result of this embedding be denoted $\tilde{\mathbf{X}}$, then we have

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_{\tau_1} \\ \vdots \\ \mathbf{X}_{\tau_{N_\tau}} \end{bmatrix} \in \mathbb{R}^{N_\mathbf{x}\cdot N_\tau \times T_\mathbf{x}}, \tag{2.30}$$

where $\mathbf{X}_{\tau_i}$ denotes the copy of $\mathbf{X}$ that is shifted by time lag τ_i. The optimal N_τ can be found using standard model selection procedures such the *Akaike Information Criterion* which was introduced for this purpose in the context of CCA (Akaike, 1976). In practice however, it is sufficient to use the well established knowledge about the neurovascular coupling dynamics to restrict the length of the temporal window to less than 20 seconds. Using the definition above we can substitute $\tilde{\mathbf{X}}$ into the original CCA objective in Eq. (2.26) from which we obtain a temporally embedded spatial filter $\mathbf{w}_\mathbf{x}(\tau)$.

The *spatio-temporal* filter $\mathbf{w}_\mathbf{x}(\tau)$ is recovered from $\tilde{\mathbf{w}}_\mathbf{x}$ as

$$\tilde{\mathbf{w}}_\mathbf{x} = \begin{bmatrix} \mathbf{w}_\mathbf{x}(\tau_1) \\ \vdots \\ \mathbf{w}_\mathbf{x}(\tau_{N_\tau}) \end{bmatrix} \in \mathbb{R}^{N_\mathbf{x}\cdot N_\tau \times 1}\ . \tag{2.31}$$

Unfortunately if there are only a limited number of samples available and at the same time the dimensionality of the data is large, then this temporal embedding will lead to very high dimensional and ill-conditioned covariance matrices. However one can apply the *kernel trick* (Müller et al., 1997) and solve the dual formulation of the problem instead (see next paragraph). This approach is proposed as temporal kernel CCA in Bießmann et al. (2009).

The essence of the kernel trick is to implicitly map the variables into a higher (possibly infinite) dimensional feature space $\mathcal{F}$ and to apply the linear machinery there. Practically, this can be achieved by substituting the linear inner product in the original formulation of the algorithm by kernel functions $k(\cdot,\cdot)$ which represent inner products in feature space

$$k(\mathbf{x},\mathbf{y}) = \langle \xi(\mathbf{x}), \xi(\mathbf{y}) \rangle_{\mathcal{F}} \; . \tag{2.32}$$

Thus the resulting algorithm can be interpreted as running the original algorithm on the (nonlinearly) mapped objects $\xi(\mathbf{x})$ and $\xi(\mathbf{y})$. The choice of the kernel largely influences the algorithm's ability to model particular types of nonlinearity. A popular kernel which works very well in practice is the Gaussian kernel $k(\mathbf{x},\mathbf{y}) = e^{-\frac{||\mathbf{x}-\mathbf{y}||^2}{2\sigma^2}}$. A very simple kernel is the linear kernel $k(\mathbf{x},\mathbf{y}) = \mathbf{x}^\top\mathbf{y}$, which is useful if the number of input dimensions is larger than the number of samples.

Many algorithms have been "kernelized" including but not limited to PCA (Schölkopf et al., 1998), CCA (Bießmann et al., 2009), and ICA (Harmeling et al., 2003). For kernel CCA one can show that the objective is of the same type as the one in Eq. (2.28), but the covariance matrices are substituted by kernel matrices which implicitly model the correlation between variables in feature space. The ij-th element of such a kernel matrix $\mathbf{K_x}$ is given by $k(\mathbf{x}_i,\mathbf{x}_j)$.

Finally we address the issue of robustifying against the tendency to overfit and the adverse impact of outliers. Overfitting occurs when the complexity of the solution is too high relative to the sample size. In other words, there is not enough data to reliably fit the complex model. One way to avoid the overfitting problem in this case is to restrict the complexity of the solution (Vapnik, 2000), e.g., by adding a *regularization term* to the objective function of the algorithm (Müller et al., 2003). One popular choice, the Tikhonov regularization term (Tikhonov and Arsenin, 1977), penalizes the complexity of the solution. Tikhonov regularization is often called L_2-norm regularization, which refers to the mechanism used to stabilize the algorithm. The key idea is to impose a penalty on the euclidean norm (i.e. the L_2-norm) of the subspace to be found. This penalty can be easily incorporated into the standard formulations of most algorithms discussed above by adding a ridge to the diagonal of the covariance matrices, hence the regression case of Tikhonov regularization is often called *ridge regression*.

In the case of CCA, the L_2 regularized version of CCA results in a generalized eigenvalue equation just like in Eq. 2.28, with the slight modification that a ridge of height $\lambda_\mathbf{x},\ \lambda_\mathbf{y}$ is added to the covariance matrices on the right hand side of the equation, such that

$$\begin{bmatrix} 0 & \mathbf{C_{xy}} \\ \mathbf{C_{yx}} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{W}_x \\ \mathbf{W}_y \end{bmatrix} = \Lambda \begin{bmatrix} \mathbf{C_x} + \mathbf{I}\kappa_\mathbf{x} & 0 \\ 0 & \mathbf{C_y} + \mathbf{I}\kappa_\mathbf{y} \end{bmatrix} \begin{bmatrix} \mathbf{W}_x \\ \mathbf{W}_y \end{bmatrix}, \tag{2.33}$$

where $\kappa_\mathbf{x},\ \kappa_\mathbf{y}$ are the regularizers for each modality, respectively. For an introduction to the relationship between standard CCA and regularized CCA see Bie and Moor (2003).

Next to these simple cases of euclidean norm constraint regularizations, there is a spectrum of other approaches to regularize the solutions of factor models. Many approaches impose a mixture of L_2 and L_1 norms, this is often referred to as *elastic net* regularization. Other

methods penalize the L_1 norm of the factor subspace or on the sources themselves. This approach is popular in the dictionary learning community, see e.g. (Mairal et al., 2010; Varoquaux et al., 2011). More sophisticated regularization schemes impose structured sparsity constraints on the solution (Jenatton et al., 2012).

Chapter 3

Shortcomings of conventional approaches

Before we are going to discuss novel algorithms to fuse multivariate oscillatory signals with other multivariate (and possibly oscillatory) signals in the subsequent chapters, we will treat a simpler case first. The simpler case consists of finding components in the multivariate oscillatory signal that exhibit a co-modulation in power with a given univariate signal. In this chapter we analyze conventional analysis approaches for assessing co-modulation between a target signal and the power dynamics of neural sources.

For this purpose, we introduce the concept of a *target variable* (in the following denoted by z), which in principle can be any scalar function of time. In the present neuroimaging context, this target variable will typically either represent a behavioral measure as the final output of the central nervous activity (e.g. reaction time, sensory detection, task rating, motor evoked potentials, etc.) or a parameter of external stimuli (e.g. when studying how amplitude modulation of neuronal oscillations correlate with stimulus properties).

3.1 Theoretical considerations

In general, we want to investigate the relation between EEG/MEG spectral power and the z variable in order to find a possible functional relationship between neuronal amplitude modulations and a behavior or stimulus. Here we discuss three standard approaches for establishing this relationship. In approach one and two, spectral power is firstly computed for each channel/sensor. A correlation to the target variable is then either assessed per channel (approach one), or using multivariate regression (approach two). In the third approach, a linear projection from sensor space into a source component space is performed first. Then bandpower is computed in this component space and correlation to the target variable is assessed per source component.

We now discuss these approaches in more detail and highlight potential drawbacks.

3.1.1 Correlation with bandpower at sensor level

In this section we derive an analytic expression for the channelwise correlations between a univariate target variable z and bandpower dynamics (approximated by changes in variance) at a single recording channel. In order to do so, we assume a simplified scenario in which

there is one source component of interest while everything else (i.e. background activity, channel noise, and so on) is subsumed by a second component. However, this does not represent a restriction because our findings generalize to more complicated cases.

Let $\mathbf{x}_k(t)$ denote the recorded EEG signal at the k-th recording channel and $s(t)$ denote the signal of a brain source of interest. All other activity is subsumed by a variable denoted with $n(t)$. According to the linear generative model for EEG/MEG/EMG recordings we have

$$x_k(t) = a \cdot s(t) + b \cdot n(t) \;, \tag{3.1}$$

for two scalar constants a and b that determine the strength and sign with which the signals $s(t)$ and $n(t)$ are expressed at the k-th recording channel. From here on we will drop the subscript k from x as well as the time index t. Let furthermore $\phi_x(e)$ denote the power of a frequency band of interest in the e-th epoch at the k-th recording channel. The power of a given frequency band is well approximated by computing the variance of the bandpass filtered signal. Let us assume, without loss of generality, that x is bandpass filtered already. Then ϕ_x is given by the epoch-wise variance of x. Thus we have

$$\begin{aligned} \phi_x(e) &= \mathrm{Var}[x](e) \\ &= a^2 \,\mathrm{Var}[s](e) + b^2 \,\mathrm{Var}[n](e) + 2ab \,\mathrm{Cov}[s,n](e) \\ &= a^2\phi_s(e) + b^2\phi_n(e) + 2ab\,\epsilon(e) \;, \end{aligned} \tag{3.2}$$

where ϕ_s and ϕ_n denote the bandpower of s and n, respectively, and the variance and covariance is computed epoch-wise, i.e. over all time samples within the e-th epoch. Furthermore we have introduced $\epsilon(e)$ as short-hand for $\mathrm{Cov}[s,n](e)$. In order to make the following derivations tractable, we will further assume that the epoch-wise covariance between s and n is negligibly small, i.e. $\epsilon(e) \approx 0$. Note however, that this assumption does not hold in general.

With these definitions we can now inspect the correlation between ϕ_x and z, which is given by

$$\mathrm{Corr}(\phi_x, z) = \frac{\mathrm{Cov}(\phi_x, z)}{\sqrt{\mathrm{Var}(\phi_x)}} = \frac{\mathrm{Cov}(a^2\phi_s, z)}{\sqrt{\mathrm{Var}(\phi_x)}} + \frac{\mathrm{Cov}(b^2\phi_n, z)}{\sqrt{\mathrm{Var}(\phi_x)}} \;, \tag{3.3}$$

where, without loss of generality, we have assumed that z has zero mean and unit variance and we have used the decomposition of ϕ_x into $a^2\phi_s + b^2\phi_n$. The variance and covariance in the above expression are computed over epochs now.

In order to resolve the denominator in the above equations, we derive an expression for $\mathrm{Var}(\phi_x)$ as function of ϕ_s and ϕ_n next:

$$\begin{aligned} \mathrm{Var}(\phi_x) &= \mathrm{Var}(a^2\phi_s) + \mathrm{Var}(b^2\phi_n) + 2\,\mathrm{Cov}(a^2\phi_s, b^2\phi_n) \\ &= a^4 \,\mathrm{Var}(\phi_s) \left(1 + \frac{b^4 \,\mathrm{Var}(\phi_n) + 2a^2b^2 \,\mathrm{Cov}(\phi_s, \phi_n)}{a^4 \,\mathrm{Var}(\phi_s)}\right) \qquad (3.4) \\ &= b^4 \,\mathrm{Var}(\phi_n) \left(1 + \frac{a^4 \,\mathrm{Var}(\phi_s) + 2a^2b^2 \,\mathrm{Cov}(\phi_s, \phi_n)}{b^4 \,\mathrm{Var}(\phi_n)}\right) \;, \qquad (3.5) \end{aligned}$$

which gives us expressions for $\mathrm{Var}(\phi_x)$ that either factor out $\mathrm{Var}(\phi_s)$ or $\mathrm{Var}(\phi_n)$.

Substituting the last two lines back into Eq. (3.3) yields an expression for the correlation between ϕ_x and z in terms of the correlations between ϕ_s and z as well as ϕ_n and z, respectively:

$$\mathrm{Corr}(\phi_x, z) = \frac{1}{q_{n/s}} \mathrm{Corr}(\phi_s, z) + \frac{1}{q_{s/n}} \mathrm{Corr}(\phi_n, z) \,, \tag{3.6}$$

where we have defined

$$q_{n/s} := \sqrt{\left(1 + \frac{b^4 \,\mathrm{Var}(\phi_n) + 2a^2b^2 \,\mathrm{Cov}(\phi_s, \phi_n)}{a^4 \,\mathrm{Var}(\phi_s)}\right)} \quad \text{and} \tag{3.7}$$

$$q_{s/n} := \sqrt{\left(1 + \frac{a^4 \,\mathrm{Var}(\phi_s) + 2a^2b^2 \,\mathrm{Cov}(\phi_s, \phi_n)}{b^4 \,\mathrm{Var}(\phi_n)}\right)} \,. \tag{3.8}$$

Now what does this tell us? Ultimately we are interested in two things: Firstly, we want to find out about the magnitude and sign of $\mathrm{Corr}(\phi_s, z)$ in order determine if there is indeed a co-modulation between z and the power of s. If this turns out to be the case, it is beneficial to know the coefficient a in the equation $x_k(t) = as(t) + bn(t)$, in order to make inference about the origin of the signal s, i.e. to localize the respective brain area. However, the above result tells us that computing channelwise bandpower correlations, i.e. $\mathrm{Corr}(\phi_x, z)$, can only give us a skewed impression of what we are really interested in, namely $\mathrm{Corr}(\phi_s, z)$. Let us make two more simplifying assumptions, namely that neither z nor ϕ_n are correlated to ϕ_n. This causes the last term in Eq. (3.6) as well as the covariance term in $q_{n/s}$ and $q_{s/n}$ to vanish. Even in this case we find that the correlation at the k-th channel is still skewed by a channel-depended multiplicative factor:

$$\mathrm{Corr}(\phi_x, z) = \frac{1}{\sqrt{\left(1 + \frac{b^4 \,\mathrm{Var}(\phi_n)}{a^4 \,\mathrm{Var}(\phi_s)}\right)}} \mathrm{Corr}(\phi_s, z) \,, \tag{3.9}$$

which in turn depends on two aspects: (i) the ratio of power modulations $\frac{\mathrm{Var}(\phi_n)}{\mathrm{Var}(\phi_s)}$, which doesn't depend on the channel index k and is thus somewhat *global*; and (ii) the ratio of the mixing coefficients $\left(\frac{b}{a}\right)^4$, which is different for each channel and thus a *local* aspect.

From these considerations we can deduce that only in extremely favorable signal-to-noise ratio (SNR) conditions, such as for example $\mathrm{Var}(\phi_s) \approx \mathrm{Var}(\phi_n)$ and $a \gg b$, we have that $\mathrm{Corr}(\phi_x, z)$ approaches $\mathrm{Corr}(\phi_s, z)$. Even in this case, however, the mixing coefficients cannot be recovered, rendering source localization of the signal $s(t)$ very difficult, if not impossible. In cases with very low SNR, such as for example $\mathrm{Var}(\phi_s) \approx \mathrm{Var}(\phi_n)$ and $a \ll b$, we have that $\mathrm{Corr}(\phi_x, z) \approx \frac{1}{b^2} \mathrm{Corr}(\phi_s, z)$. In this case, even if there is a large correlation between ϕ_s and z, it will be strongly diminished due to the large discounting influence of background activity.

This illustrates the major drawbacks of the univariate approach:

- **Low signal-to-noise ratio**. The true correlation between the power of s and z may be heavily underestimated in the presence of strong noise sources with high power.

- **Lack of interpretability**. The noise contribution may differ across electrodes, introducing a channel-specific bias in the correlation coefficients. Hence, topographic maps of sensor-space correlation may not give a good indication of where the signal-of-interest is strongest and are thereby hard to interpret in neurophysiological terms.
- **Disregard of the generative model.** It is practically impossible to disentangle signal and noise contributions by looking only at single electrodes. More generally, this approach does not provide a factorization of the measurement $\mathbf{x}(t)$ into $\mathbf{A}$ and $\mathbf{s}(t)$ according to the generative linear model of EEG/MEG data (see Eq. (2.1)).

Note that such a (successful) factorization would deliver signal and noise components in different rows of $\mathbf{s}(t)$ and their spatial field patterns (which are interpretable in terms of the spatial locations of the components) in the corresponding columns of $\mathbf{A}$, thereby effectively circumventing the aforementioned problems.

3.1.2 Correlation with a linear combination of sensor-level bandpower

Instead of correlating the bandpower of individual channels with z, multivariate approaches can be applied in order to combine the bandpower signals from all channels before correlating the resulting bandpower time-course with the target variable z. The weights for a linear combination of channelwise bandpower time-courses can be optimized by regression, thereby effectively maximizing the correlation.

Let us extend our simple model from the previous section to the multivariate case, i.e. let

$$\mathbf{x}(t) = \mathbf{a} \cdot s(t) + \mathbf{b} \cdot n(t) \,, \tag{3.10}$$

where the vector $\mathbf{x}(t) \in \mathbb{R}^{N_x}$ represents the recorded signals of N_x channels at time point t and the vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N_x$ represent the mixing coefficients for the signal of interest $s(t)$ and the background signal $n(t)$, respectively.

Using the same formalism we have seen above, we define the now multivariate sensor-space bandpower signal $\phi_{\mathbf{x}}$ is given by

$$\begin{aligned} \phi_{\mathbf{x}}(e) &= \mathrm{Var}[\mathbf{x}](e) \\ &= \mathbf{a} \circ \mathbf{a} \cdot \mathrm{Var}[s](e) + \mathbf{b} \circ \mathbf{b} \cdot \mathrm{Var}[n](e) + 2 \cdot \mathbf{a} \circ \mathbf{b} \cdot \mathrm{Cov}[s, n](e) \\ &= \mathbf{a} \circ \mathbf{a} \cdot \phi_s(e) + \mathbf{b} \circ \mathbf{b} \cdot \phi_n(e) + 2 \cdot \mathbf{a} \circ \mathbf{b} \cdot \epsilon(e) \,, \end{aligned} \tag{3.11}$$

where $\mathrm{Var}[\mathbf{x}](e)$ denotes the vector of channel-wise variance computed within the e-th epoch. Furthermore, $(\mathbf{a} \circ \mathbf{b}) \in \mathbb{R}^{N_x}$ denotes the element-wise multiplication of the vectors $\mathbf{a}$ and $\mathbf{b}$. Accordingly $\mathbf{a} \circ \mathbf{a}$ denotes the element-wise squaring of the vector $\mathbf{a}$ and similarly for $\mathbf{b}$. The rest is the same as in the previous section.

The question is now if it is possible to find a weight vector $\mathbf{w}$ such that $\mathbf{w}^\top \phi_{\mathbf{x}}(e)$ closely resembles $\phi_s(e)$. From the last equation it can be seen that ϕ_s can only be fully recovered from the multivariate power signal $\phi_{\mathbf{x}}$ if the weight vector $\mathbf{w}$ is orthogonal to the vector $\mathbf{b} \circ \mathbf{b}$ *and* to the vector $\mathbf{a} \circ \mathbf{b}$, i.e. such that $\mathbf{w}^\top (\mathbf{b} \circ \mathbf{b}) \approx \mathbf{w}^\top (\mathbf{a} \circ \mathbf{b}) \approx 0$. Only in this case we have $\mathbf{w}^\top \phi_{\mathbf{x}}(e) \approx c \cdot \phi_s(e)$, where the constant c is the scalar product between the vectors $\mathbf{w}$ and $\mathbf{a} \circ \mathbf{a}$.

The question is of course how to find such $\mathbf{w}$ if $\mathbf{a}$ and $\mathbf{b}$ are unknown. If it indeed holds that $z \approx \phi_s$, then regression using z as the target signal will lead to a $\mathbf{w}$ that maximizes the correlation between $\mathbf{w}^\top \phi_{\mathbf{x}}$ and ϕ_s. Thus, this multivariate approach may achieve superior performance compared to the univariate correlation approach outlined above.

However, in principle the other two drawbacks remain:

- **Disregard of the generative model**. Linearly regressing power values violates the underlying generative model for EEG and MEG data, because the generative model is linear in the 'raw' data, not in the power transformed, i.e. squared, data.
- **Lack of interpretability**. The regression weight vector does not necessarily contain neurophysiologically interpretable information about the location of the underlying correlating source, because it is again a function of both the signal-of-interest and the noise. Even if the corresponding spatial pattern is derived from the weight vector, it will be defined in the power transformed data space and can thus not be subjected to source-localization algorithms.

3.1.3 Correlation with bandpower at source level - spatial filtering approaches

We have seen that even if the power time-course of a source of interest can be recovered using a multivariate approach, interpretability of the resulting model (i.e. the weight vector or the corresponding spatial pattern) with respect to the origin of the source is low. The reason for this shortcoming of the previous two approaches is that the bandpower is computed at the level of sensors, which corresponds to a *non-linear transformation* of the *linearly* mixed signals. Thus, a linear unmixing of the multivariate bandpower signal can, in general, not recover the original mixing model. This, however, is required in the context of neuroimaging data in order to allow for neurophysiological interpretation.

The solution to the problems of the previously discussed approaches is thus to change the order of operations, i.e. to first apply a linear unmixing and then apply the non-linearity, that is compute bandpower. Linear unmixing amounts to projecting the data into a hypothetical source (or component) space. Such a transformation into source space can be obtained by way of so-called spatial filtering methods. These methods estimate a linear projection of the data $\mathbf{x}(t)$ onto a set of weight vectors, which are represented here by the matrix $\mathbf{W} \in \mathbb{R}^{N_x \times K}$, where $\mathbf{W} = \left(\mathbf{w}^1, \ldots, \mathbf{w}^i, \ldots, \mathbf{w}^K\right)$, $\mathbf{w}^i \in \mathbb{R}^{N_x}$, and K corresponds to the number of estimated sources $\hat{s}_i(t)$. In accordance with the literature, we refer to these weight vectors $\mathbf{w}^i$ as *spatial filters*. Each of the spatial filters is meant to extract the signal from one source while suppressing the activity of the others, such that the resulting projected signal is a close approximation of the original source signal, i.e. $\hat{s}_i(t) = {\mathbf{w}^i}^\top \mathbf{x}(t)$. Bandpower correlations are then computed using $\hat{s}_i(t)$ instead of $x_k(t)$.

Popular spatial filtering approaches, such as ICA, are in line with the generative model of EEG/MEG and can thus in principle deliver results that are interpretable within this model. Please note, however, that approaches such as ICA are "blind" by definition, i.e. they base the decomposition into source signals solely on assumptions about the statistics of the estimated sources. This is the reason why this approach is often referred to as *blind source*

separation (BSS). In our setting, where additional information is available in the form of the target variable z, it would be entirely ignored by the BSS method. Furthermore, in ICA there is no natural ordering of the estimated components and the number of components that are estimated by the method may be higher than one is actually interested in.

Taken together, we find that BSS approaches can yield results that are interpretable within the generative model of EEG/MEG but the ignorance of the available target function may lead to suboptimal performance compared to supervised approaches.

The order of operations is important. The main difference between the last two multivariate approaches to uncovering ϕ_s is the order in which the following mathematical operations are applied: (i) computation of bandpower and (ii) linear combination, i.e. application of a weight vector $\mathbf{w}$. Let us use the symbol $\phi(\cdot)$ to denote the computation of bandpower. Then we can express the application of the aforementioned operations to the data $\mathbf{x}$ in different order by $\mathbf{w}^\top\phi(\mathbf{x})$ and $\phi(\mathbf{w}^\top\mathbf{x})$, respectively. Note that the result of $\mathbf{w}^\top\phi(\mathbf{x})$ is in general different from $\phi(\mathbf{w}^\top\mathbf{x})$ because computing bandpower and applying a linear combination are nonlinear and linear operations, respectively. Therefore, they are not interchangeable. Mathematically speaking, these two operations do not commute.

We illustrate the difference between $\mathbf{w}^\top\phi(\mathbf{x})$ and $\phi(\mathbf{w}^\top\mathbf{x})$ in Figure 3.1. The figure shows a two dimensional example, similar to what was proposed in the theoretical considerations above. There is one source of interest (red) and one distractor source (blue) in the unobservable source space. These signals are mapped to the sensor space by way of Eq. (2.1). The target variable z corresponds to the bandpower dynamics of the source of interest. Computing bandpower of the sensor space signals and then applying a linear combination (i.e. $\mathbf{w}^\top\phi(\mathbf{x})$) leads to only a crude approximation of the target variable. However, applying a linear combination of the sensor space signals and then computing bandpower of the resulting signal (i.e. $\phi(\mathbf{w}^\top\mathbf{x})$) recovers the original source time course and therefore also its bandpower modulation.

A further illustration is given in Figure 3.2. In this figure, the sensor space data are plotted as a scatter plot. This way, we can show the columns of the mixing matrix (in this case the vectors $\mathbf{a}$ and $\mathbf{b}$) as well as the weight vector $\mathbf{w}$ that maximizes the correlation between the target variable and the estimated bandpower time-course. Since it is only a two-dimensional example, the weight vector $\mathbf{w}$ can be parameterized by a single number, namely its angle with respect to one of the axis of the coordinate system. Thus it is possible exhaustively search in this one-dimensional space and visualize the bandpower correlation as a function of the weight vector. In Figure 3.2 this is done for three different configurations of the mixing matrix. In fact, we change only one column of the mixing matrix, namely the *spatial pattern* that corresponds to the target source, i.e. the vector $\mathbf{a}$. Furthermore, the figure shows the data after channelwise computation of bandpower (and subsequent centering), along with the vectors $\mathbf{a} \circ \mathbf{a}$, $\mathbf{b} \circ \mathbf{b}$, and $\mathbf{a} \circ \mathbf{b}$ as well as the weight vector $\mathbf{w}$ that maximizes the correlation between $\mathbf{w}^\top\phi(\mathbf{x})$ and z. Recall that the vector $\mathbf{a} \circ \mathbf{b}$ was introduced due to cross-terms that arise due to the squaring of the mixing process. It can be seen that for some scenarios (top and middle row in the figure), the approach $\mathbf{w}^\top\phi(\mathbf{x})$ yields satisfactory correlations with z. However, the quality of the solution greatly depends on the configuration of the vectors $\mathbf{a}$ an $\mathbf{b}$ and on the resulting mix-term vector $\mathbf{a} \circ \mathbf{b}$. Note however, that the

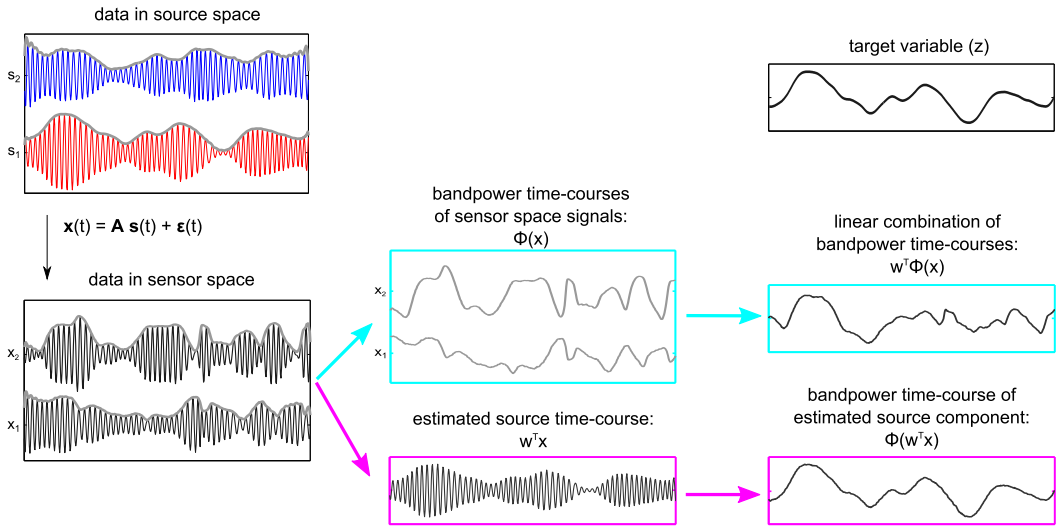


Figure 3.1: The difference between $\mathbf{w}^\top \phi(\mathbf{x})$ and $\phi(\mathbf{w}^\top \mathbf{x})$, i.e. applying nonlinear and linear operations in different order. See main text for detailed explanation.

spatial filtering approach $\phi(\mathbf{w}^\top \mathbf{x})$ is entirely unaffected by this and is therefore able to recover the true bandpower dynamics perfectly.

3.2 Simulations

After the theoretical considerations of the previous section, we now describe a number of simulations that quantify some of the concerns that were raised above. Specifically we evaluate the discussed approaches with respect to two important aspects: (i) how well they extract a power modulation signal of interest, especially on data that was not used to fit the model, and (ii) to what degree the model parameters can be evaluated with respect to the origin of the extracted signal. Let us denote these two aspects by *accuracy* and by *interpretability*, respectively.

3.2.1 Simulation setup

Data generation

Simulated 58 channel EEG data was created according to the generative model outlined in section 2.2.1. The mixing matrix $\mathbf{A}$ and the source time courses $\mathbf{s}$ were created separately.

Firstly, we generated time courses of $N_{\text{bg}} + 1$ hypothetical band-limited EEG sources (1 target source and $N_{\text{bg}} = 100$ background sources). For illustrative purposes we chose the α-band as the frequency band of interest, i.e. 8 to 12 Hz. The oscillatory signals were created individually by separately constructing the amplitude and phase spectrum and

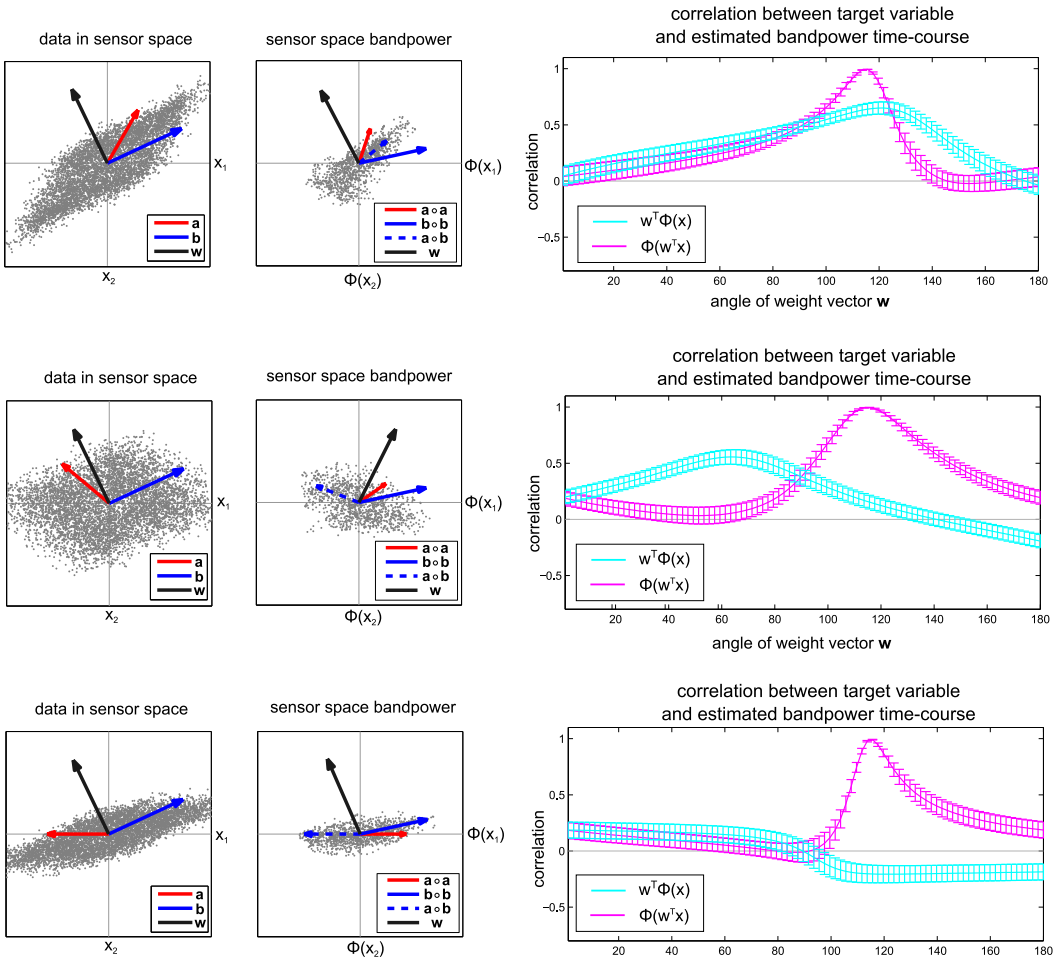


Figure 3.2: A further illustration of the difference between $\mathbf{w}^\top \phi(\mathbf{x})$ and $\phi(\mathbf{w}^\top \mathbf{x})$, i.e. applying nonlinear and linear operations in different order. See main text for detailed explanation.

then using inverse Fourier transform to obtain the time-domain signal. In the amplitude spectrum, the coefficients of the alpha band were set to 1, whereas the amplitude of all other frequencies were set to zero. The phase spectrum was chosen randomly for each source time course. Once the time-domain source signals were constructed, their envelopes were normalized to 1. Thereafter the signals were multiplied with an amplitude modulation function that consisted of low-pass filtered white noise (filter cut-off below 0.5 Hz). An offset was added such that the slow amplitude modulation was always larger than zero. Squaring the amplitude modulation function of a source yields the power modulation of that source. This constitutes the EEG data in 'source space'.

Physiologically plausible spatial patterns were generated via a realistic EEG forward model (Nolte and Dassios, 2005; Fonov et al., 2011). Specifically, we placed model neural sources (i.e. electrical dipoles, here with randomly chosen orientation) at randomly chosen locations in 3D voxel space and computed the resulting EEG sensor space projections, which we denote with the vector $\mathbf{a}^i \in \mathbb{R}^{N_x}$ for the i-th source, where $N_x = 58$ denotes the number of simulated EEG channels. See Haufe et al. (2012) for a related simulation setup.

The following equations describe the generation of the pseudo-EEG measurement $\mathbf{x}(t)$ as the sum of contributions $\mathbf{x}_\mathrm{s}(t)$ due to target sources, contributions $\mathbf{x}_\mathrm{ba}(t)$ due to background sources, and sensor noise $\mathbf{x}_\mathrm{noise}(t)$:

$$\mathbf{x}_\mathrm{s}(t) = \mathbf{a}_\mathbf{x}^1 \, s_1(t) \tag{3.12}$$

$$\mathbf{x}_\mathrm{ba}(t) = \sum_{i=1}^{N_\mathrm{bg}} \mathbf{a}_\mathbf{x}^{i+1} \, s_{i+1}(t) \tag{3.13}$$

$$\mathbf{x}_\mathrm{noise}(t) = \frac{1}{c_\mathrm{ba}} \, \mathbf{x}_\mathrm{ba}(t) + \gamma_\epsilon \, \epsilon_\mathbf{x}(t) \tag{3.14}$$

$$\mathbf{x}(t) = \gamma \frac{\mathbf{x}_\mathrm{s}(t)}{c_\mathrm{s}} + \frac{\mathbf{x}_\mathrm{noise}(t)}{c_\mathrm{noise}} \, . \tag{3.15}$$

The constants c_ba, c_s, and c_noise are the norm constants of the corresponding data matrices $\mathbf{x}_\mathrm{ba}$, $\mathbf{x}_\mathrm{s}$, and $\mathbf{x}_\mathrm{noise}$, and serve to balance the data matrices prior to building the weighted sums. For the matrix $\mathbf{x}_{(\cdot)}$, the corresponding Frobenius norm is given by

$$c_{(\cdot)} = \sqrt{\sum_i^{N_x} \sum_t^{T} \left(\mathbf{x}_{(\cdot)}(t)\right)_i^2} \, , \tag{3.16}$$

i. e., it measures the signal energy, averaged across channels and time. The scaling parameter γ_ϵ determines the strength of (Frobenius-) normalized Gaussian noise $\epsilon_\mathbf{x}$, and was set to 0.1 in all simulations. The scaling parameter $\gamma > 0$ determines the ratio between the energy of the scalp-projected target source time courses and the energy of all other (background and sensor noise) contributions in the data. Thus the signal-to-noise ratio in dB scale is given by

$$\mathrm{SNR} \quad = \quad 20 \log_{10}(\gamma) \, . \tag{3.17}$$

Data analysis

Target function In this simulation the target function z was the true envelope of the target source, i.e. the source to be extracted. This means that the correlation between the target function z and the bandpower dynamics of the estimated target source could reach up to it's maximum, namely one, if the estimated target source corresponded to the true target source.

Compared models We compared three multivariate approaches to extract a bandpower signal that co-modulates with the given target function, namely regression, PCA, and ICA, which all have been introduced in chapter 2.3. Additionally we used an "oracle method" in order to map out an empirical upper bound of performance. This oracle method consisted of an ordinary-least-squares (OLS), which optimizes a spatial filter using the true source component signal. Since this is not possible outside of simulations in which the time-course of the true source component is known, we consider it an oracle method that helps to give some indication of how much the compared approaches deviate from an empirical optimum.

Note that regression is supervized and makes use of the target variable z. Furthermore, regression is applied to channel-wise computed bandpower as outlined above, while PCA and ICA are applied to the time-domain data and bandpower is computed on the time-courses of the PCA/ICA components. Epoch length is set to 500 ms. While regression only estimates a single source, PCA and ICA return as many components as there are channels. Thus, for PCA and ICA, the putative target source was determined based on bandpower correlations of all components to the target function, obtained on training data.

Metrics

The methods were compared with respect to (i) how well they recover the bandpower dynamics of the target source (*accuracy*) and (ii) how well the resulting model parameters can be interpreted as to the neural origin of the estimated source (*interpretability*). Note that all metrics are computed on test data, i.e. data that was not used to train the models.

Accuracy The estimated bandpower time-course was correlated to the known bandpower time-course of the target source.

Interpretability All methods are backward models and as such they optimize weight vectors to extract the signal of interest from the data. Spatial activation patterns were computed from the weight vectors of the models according to the filter/pattern transformation introduced in chapter 2.2, section 2.2.3. Note that for regression, the spatial pattern is "power pattern" only, because regression is applied to the non-linearly transformed data. The resulting patterns of the methods were correlated with the true sensor space spatial activation pattern of the target source. A second interpretability metric is derived from attempting to source-localize the resulting sensor space patterns and then computing the distance to the true

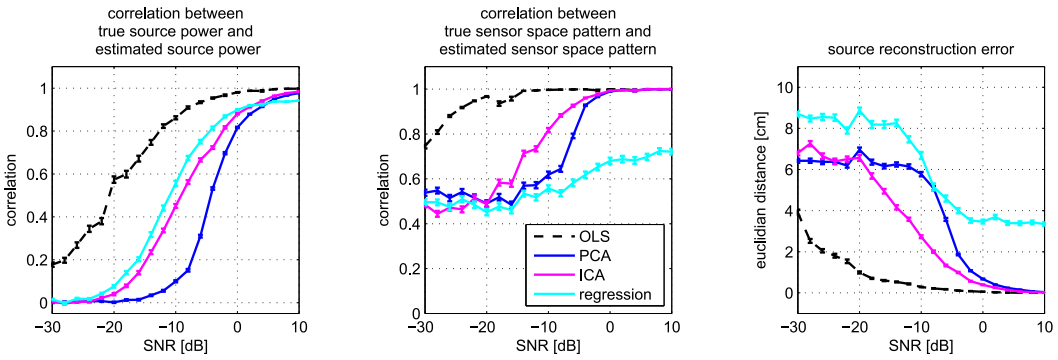


Figure 3.3: Results of simulation that compares regression, PCA, and ICA in the task to extract the power dynamics of source signal of interest. OLS is an oracle method that shows how well the target signal can be extracted from the data if it is already known and thus, represents an empirical optimum to benchmark the other approaches against. The plots indicate the performance of the compared methods in terms of *accuracy* and *interpretability*.

source location. The source localization error was assessed by first fitting a dipole to the estimated sensor-space pattern using the MUSIC algorithm (Schmidt, 1986). Thereafter the Euclidean distance to the true dipole of the target source was computed.

3.2.2 Results

The simulations were repeated 100 times for a range of SNR settings. Figure 3.3 displays the performance in terms of the accuracy and interpretability measures as a function of SNR.

Being a supervised model, regression on bandpower representations yields good accuracy. Over a large range of SNR settings it yields the best approximation of the true source power dynamics among the compared methods. It can also be seen that ICA yields much better accuracy in comparison to PCA, i.e. it is able to invert the generative model to a much better degree.

However, regression does not perform very well in terms of interpretability, which is due to it not being in line with the generative model of EEG/MEG as outline above. Source separation methods such as PCA and ICA on the other hand, respect the linear generative model and thus deliver better interpretability.

3.3 Conclusion

We find that the simulations demonstrate the concerns that we made based on theoretical considerations earlier in this chapter.

Accuracy and interpretability are two highly desirable properties of neuroimaging models and thus models should be evaluated with respect to both. Computing band-power at the level of recording channels and then applying linear methods to the resulting time-frequency representation is not in line with a widely accepted generative model of EEG/MEG, thus hindering the neurophysiological interpretation of the resulting model parameters.

In fact, none of the compared approaches exceeded in both accuracy and interpretability, thus indicating a potential gain in combining the advantages of the approaches. These advantages are adherence to the generative model in the case of ICA and use of additional information provided by the target variable in the case of regression. In the multimodal setting, the latter aspect translates into taking information from the other modality into account.

Chapter 4

Optimizing Source Power Co-modulation

In the previous chapter we have seen that blind source separation (BSS) approaches are in principle in line with the generative model. However, they do not make us of the target variable and are thus potentially suboptimal. In order to alleviate these shortcomings, we present a novel spatial filtering approach, specifically designed for relating source power modulations to a known target signal. We call this supervised approach Source Power Co-modulation (SPoC).

The core idea of the SPoC approach is to (i) decompose the multivariate EEG/MEG data into a set of source components and (ii) to use the information contained in the target variable to guide the decomposition. The result of this approach is a set of spatial filters, $\mathbf{W}$, which *directly* optimize the co-modulation between the target variable z and the power time course of the spatially filtered signal. Fig. 4.1 illustrates the contrast between regression of channel-wise band power features, BSS methods, and the SPoC approach.

In the following subsection, we describe two algorithms that implement the SPoC approach. We refer to these two methods as SPoC_{r2} and SPoC_{λ}, respectively. The difference between SPoC_{r2} and SPoC_{λ} lies in the exact definition of co-modulation between band power and z: SPoC_{r2} optimizes *correlation*, while SPoC_{λ} optimizes *covariance*. However, both algorithms invert the generative model given in Eq. (2.1) *prior* to the computation of band power and thereby avoid pitfalls that were outlined in the previous chapter.

4.1 SPoC algorithms: SPoC_{λ} and SPoC_{r2}

4.1.1 Assumptions and definitions

We assume that the EEG/MEG data $\mathbf{x}(t)$ has been band-pass filtered in the frequency band of interest. Thus, the power of the projected signal $\mathbf{w}^\top\mathbf{x}(t)$ within a given time interval is well approximated by the variance of $\mathbf{w}^\top\mathbf{x}(t)$ within that interval. We refer to such time intervals as *epochs* and assume that the EEG/MEG data can be divided up into consecutive or overlapping epochs of suitable length[1]. Epochs will be indexed by the index e.

[1] Working with epoched data instead of continuous data does not represent a loss of generality, because all of the following derivations can be reformulated for continuous data as well, provided that the target variable changes slowly enough. We choose to work with epoched data because it resembles the format of data obtained in trial-based experiments.

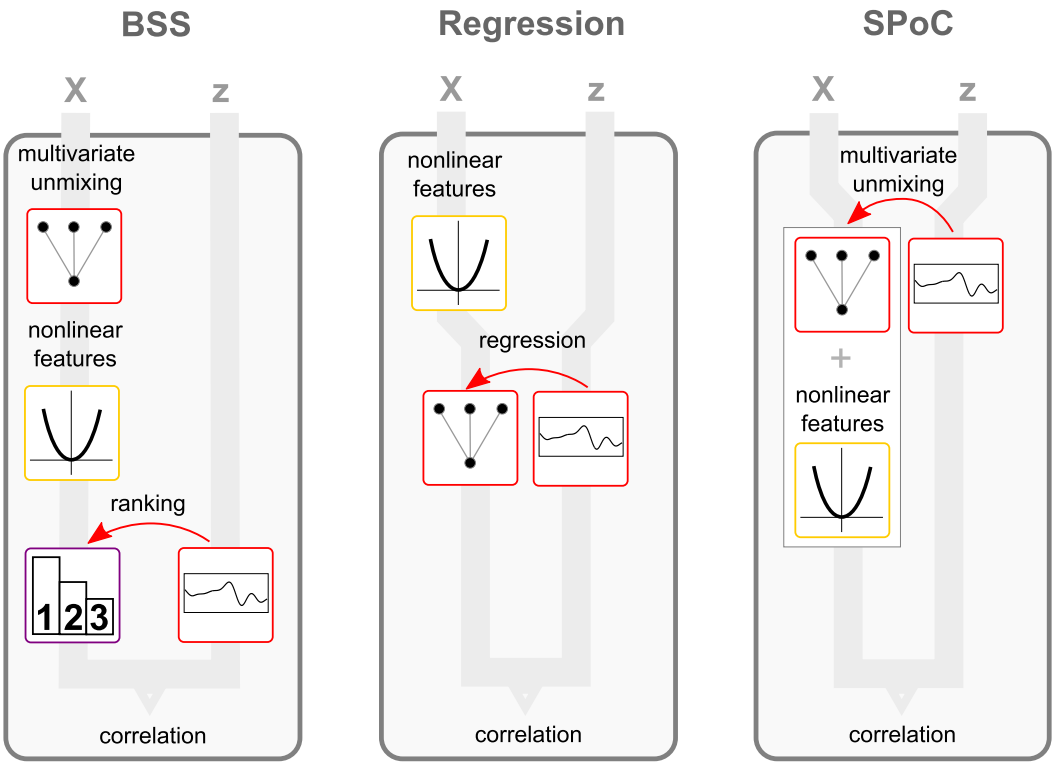


Figure 4.1: Illustration of different approaches to relating spectral power to a target variable. The input to all three approaches is the multivariate and bandpass filtered EEG/MEG data $\mathbf{x}(t)$ as well as the target variable z. Processing steps are organized from top to bottom. **Left:** An approach that is based on blind source separation (BSS) methods. A BSS method such as ICA tries to estimate the sources prior to computation of spectral power. This approach is in line with the generative EEG model and in principle could have the potential to find the true source. However, BSS techniques do not make use of the information contained in z and is bound to fail for low SNR or if the number of sources is larger than the number of channels. **Middle:** An approach that is based on regression. First, spectral power is computed on each sensor. Then the power time courses are linearly combined to resemble z as close as possible. **Right:** Our novel SPoC approach method makes use of z to guide the source estimation and to give preference to sources whose power time course resembles z. Spectral power is computed on the estimated *source components*.

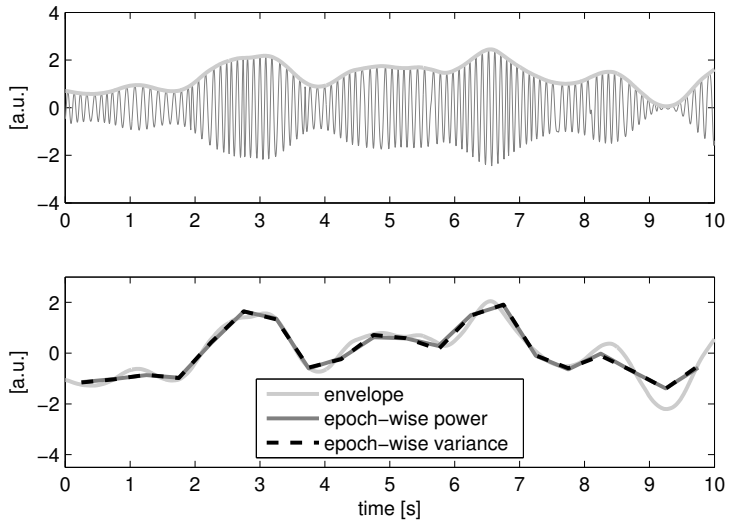


Figure 4.2: Approximation of amplitude modulations by epoch-wise bandpower and epoch-wise variance. Top: A ten seconds excerpt of a band-limited (8 Hz to 12 Hz) oscillatory signal that exhibits ongoing amplitude modulations. The amplitude modulations are also called the *envelope* (thick gray line) of the signal. Bottom: The envelope of the above signal (normalized to zero mean and unit standard deviation) together with two approximations given by the epoch-wise power (squared envelope, averaged within epochs) and the epoch-wise variance.

We assume the target variable z to only have a single value per epoch, which can be achieved by appropriate resampling. Furthermore we assume without loss of generality that z has zero mean and unit variance, which can be achieved by normalization.

It is our goal to approximate the target variable z with the bandpower/variance of a source component. We denote this estimate by ϕ, which depends on a spatial filter $\mathbf{w}$. Let $\mathrm{Var}\left[\mathbf{w}^\top \mathbf{x}(t)\right](e)$ denote the variance of $\mathbf{w}^\top \mathbf{x}(t)$ in a given epoch e. This epoch-wise variance of the projected signal will serve as the approximation of z. Thus we have

$$z(e) \approx \phi(e) = \mathrm{Var}\left[\mathbf{w}^\top \mathbf{x}(t)\right](e) = \mathbf{w}^\top \mathbf{C}(e)\mathbf{w}\ , \tag{4.1}$$

where $\mathbf{C}(e)$ denotes the covariance matrix of the e-th epoch. Figure 4.2 shows an example of approximating ongoing bandpower modulations by epoch-wise variance.

Accordingly, the first two moments of ϕ can be expressed in terms of the weight vector and epoch-wise covariance matrices. The mean of ϕ over epochs is given by

$$\begin{aligned} \langle \phi(e) \rangle &= \left\langle \mathbf{w}^\top \mathbf{C}(e)\mathbf{w} \right\rangle \\ &= \mathbf{w}^\top \left\langle \mathbf{C}(e) \right\rangle \mathbf{w} \\ &= \mathbf{w}^\top \mathbf{C}\mathbf{w}\ , \end{aligned} \tag{4.2}$$

where $\langle \cdot \rangle$ denotes the average across epochs and $\langle \mathbf{C}(e) \rangle = \mathbf{C}$ is the average covariance matrix.

For the variance of ϕ we have

$$\begin{aligned}\mathrm{Var}\left[\phi(e)\right] &= \left\langle \phi(e)^2 \right\rangle - \left\langle \phi(e) \right\rangle^2 \\ &\quad \left\langle \left(\mathbf{w}^\top \mathbf{C}(e)\mathbf{w}\right)^2 \right\rangle - \left(\mathbf{w}^\top \mathbf{C}\mathbf{w}\right)^2 \ . \end{aligned} \tag{4.3}$$

Finally, we define the matrix

$$\mathbf{C}_z := \left\langle \mathbf{C}(e) z(e) \right\rangle , \tag{4.4}$$

which helps to conveniently express the covariance between ϕ and z as

$$\begin{aligned}\mathrm{Cov}\left[\phi(e), z(e)\right] &= \left\langle \phi(e) z(e) \right\rangle - \left\langle \phi(e) \right\rangle \left\langle z(e) \right\rangle \\ &= \left\langle \left(\mathbf{w}^\top \mathbf{C}(e)\mathbf{w}\right) z(e) \right\rangle \\ &= \mathbf{w}^\top \left\langle \mathbf{C}(e) z(e) \right\rangle \mathbf{w} \\ &= \mathbf{w}^\top \mathbf{C}_z \mathbf{w}, \end{aligned} \tag{4.5}$$

where in the transition from line 1 to line 2 of this equation we have used the fact that $\langle z(e) \rangle = 0$.

Equipped with these definitions we now formulate the objectives optimized by the two algorithms that implement the SPoC approach. For ease of notation, the derivations are given for a single weight vector $\mathbf{w} \in \mathbb{R}^{N_x}$, but generalize naturally to multiple filters providing a full-rank decompositions of the data matrix.

4.1.2 SPoC$_\lambda$

Firstly, we show how to optimize the *covariance* between ϕ and z. As we will see shortly, this leads to an objective function that has a number of computationally desirably properties. We refer to this algorithm as SPoC$_\lambda$. Unlike the correlation, the covariance is affected by the scaling of its arguments. Thus far we have assumed z to have zero mean and unit variance, i.e. that the scaling of z is limited. However, it is furthermore necessary to limit the scaling of ϕ in order to prevent the covariance objective from diverging. Thus, we impose a constraint on the norm of $\mathbf{w}$ and thereby limit the scaling of ϕ. Specifically we choose the constraint such that the output of the spatial filter has unit variance.

With these definitions we arrive at the following objective function:

$$f_\lambda = \mathrm{Cov}\left[\phi(e), z(e)\right] = \mathbf{w}^\top \mathbf{C}_z \mathbf{w}, \tag{4.6}$$

with respect to the following norm constraint:

$$\mathrm{Var}\left[\mathbf{w}^\top \mathbf{x}(t)\right] = \mathbf{w}^\top \mathbf{C}\mathbf{w} \stackrel{!}{=} 1. \tag{4.7}$$

This constraint optimization problem can be solved using the method of Lagrange multipliers. Setting the first derivative of the corresponding Lagrangian to zero leads to the following generalized eigenvalue equation:

$$\mathbf{C}_z \mathbf{w} = \lambda \mathbf{C}\mathbf{w}, \tag{4.8}$$

where the eigenvalue λ corresponds to the value of f_λ evaluated at the respective eigenvector $\mathbf{w}$. Thus λ can directly be interpreted as the covariance between ϕ and z.

Finding the solution to optimizing f_λ is not as time consuming as iterative optimization procedures and the obtained solution is unique, i.e. no restarts are necessary. Furthermore, the result of solving the generalized eigenvalue problem is a full set of weight vectors, i.e. a matrix $\mathbf{W}$ with the eigenvectors in its columns. This matrix contains a column vector $\mathbf{w}$ that *maximizes* f_λ as well as a different column $\mathbf{w}$ that *minimizes* the same objective function. After sorting the columns of $\mathbf{W}$ according to their respective eigenvalues (i.e. covariances), one finds these weight vectors in the first and last column of $\mathbf{W}$. The matrix $\mathbf{W}$ has full rank but its columns are *not* mutually orthogonal, as is the case in PCA for example.

4.1.3 SPoC$_{r2}$

In this subsection we show how to optimize the *correlation* between ϕ and z. We are interested in positive as well as negative correlations and hence choose to maximize the squared correlation between z and ϕ. We refer to this SPoC algorithm as SPoC$_{\text{r2}}$ and maximize the following objective function:

$$\begin{aligned} f_{\text{r2}} &= \text{Corr}\left[\phi(e), z(e)\right]^2 \\ &= \frac{\text{Cov}\left[\phi(e), z(e)\right]^2}{\text{Var}[\phi(e)]\,\text{Var}[z(e)]} \\ &= \frac{\left(\mathbf{w}^\top \mathbf{C}_z \mathbf{w}\right)^2}{\left\langle \left(\mathbf{w}^\top \mathbf{C}(e) \mathbf{w}\right)^2 \right\rangle - \left(\mathbf{w}^\top \mathbf{C} \mathbf{w}\right)^2}. \end{aligned} \tag{4.9}$$

In the last equality of Eq. (4.9) we have used Eq. (4.5), Eq. (4.3), and the fact that $\text{Var}[z(e)] = 1$.

The weight vector $\mathbf{w}$ that maximizes f_{r2} cannot be found analytically. It should therefore be found using iterative optimization methods such as gradient descent for example. We propose to optimize it by means of standard nonlinear optimization techniques such as the limited-memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) algorithm (Nocedal, 1980) implemented in MATLAB's (The Mathworks) `fminunc` routine. After the first $\mathbf{w}$ is found, subsequent components can be extracted using the deflation scheme outlined in chapter A.3.

4.2 Validation

SPoC is designed to find a (set of) spatial filter(s) that extract oscillatory signals with power modulations that follow a given target variable. We test this ability in high dimensional and noisy environments by applying SPoC as well as linear regression and a BSS method (here we used ICA) to simulated as well as real EEG data.

In the simulations, the time course of the source signal (and therefore also its power modulation) are known. Thus the results of the methods can be compared to the ground truth. For the validation using real EEG data, we conducted an auditory steady state experiment in which the near linear relationship between stimulus intensity and neuronal amplitude modulations have been reported before (Picton et al., 2003).

4.2.1 Simulations

The data generation processes and the evaluation metrics are identical to what was described in chapter 3.2. The methods are compared in terms of *accuracy* and *interpretability*, where accuracy describes how well the bandpower time-course of the target source component is extracted and interpretability reflects the similarity of the estimated activation pattern to that of the true target source.

In addition to varying the signal-to-noise ratio between the target source and the background sources, we also explored the performance with respect to varying amounts of training data available and with respect to how well the target variable z (which is used by the supervised algorithms) reflects the bandpower time-course of the true target source. The latter aspect pertains to how the supervised algorithms perform once their training signal no longer reflects the signal that is to be extracted.

In the simulation runs in which the SNR was varied, there we two minutes of training data and two minutes of test data. In runs in which the amount of training was varied, the SNR was set to -10 dB.

Data analysis

The application of regression and ICA was conducted as described in chapter 3.2. For SPoC_λ, the resulting components were ordered with respect to the respective eigenvalues. After ordering the SPoC component set obtained from the training data, the first component was used for evaluation on the test data. For SPoC_{r2} only one weight vector was obtained by optimizing the respective objective functions on the training data.

Results

Fig. 4.3 shows the results of the simulations in which the signal-to-noise ratio (SNR) was varied. Please note that all reported correlations are obtained on test data that was not used to train the algorithms. It can be seen, that for higher SNRs all methods have satisfactory performance, i.e. they are able to extract the target power time-course from the data with a high degree of correlation. In terms of pattern similarity and source reconstruction error, PCA, ICA and the SPoC algorithms reach near perfect performance at high SNR regimes. As we have seen in the previous chapter, this is not the case for the spatial patterns obtained from regression.

However, in lower SNR regimes the SPoC algorithms clearly outperform ICA and regression. Comparing the performance of the SPoC methods, we find that the performance of SPoC_{r2}

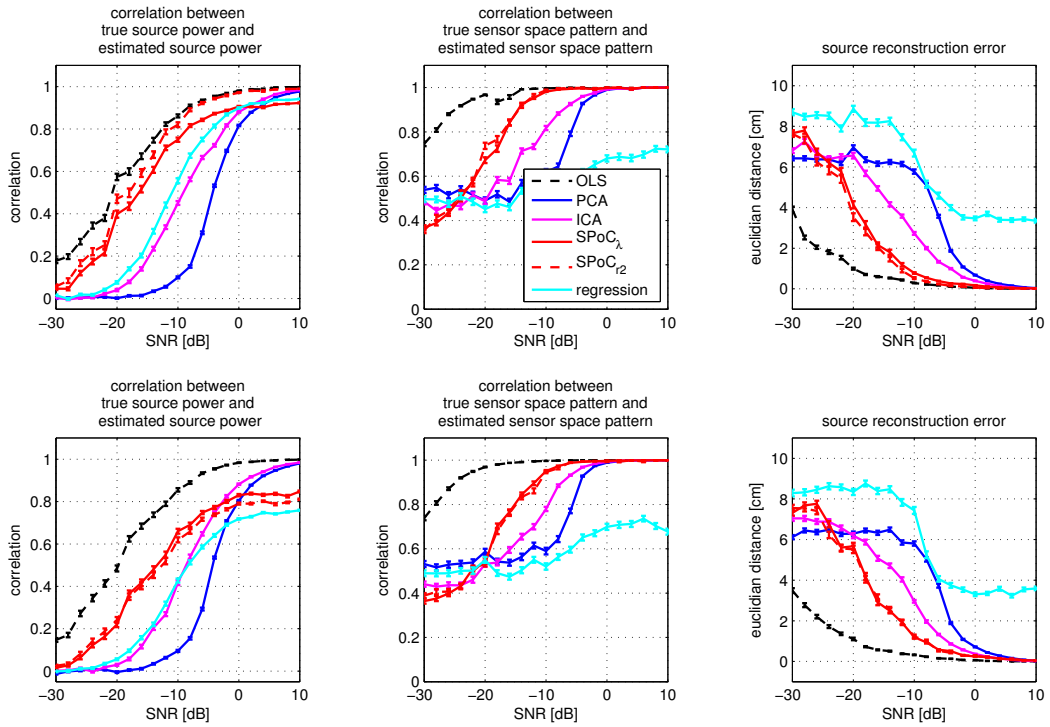


Figure 4.3: Simulation results. Here performance in terms of accuracy and interpretability is shown as a function of signal-to-noise ratio. In the top row of the plots, the target variable z (which is used for training the SPoC algorithms and regression) and the bandpower of the target source were set to correlate with each other perfectly. In the simulations depicted in the bottom row, that correlation was set to 0.7.

is consistently higher than the performance of SPoC$_\lambda$. This does not come as a surprise because only SPoC$_{\text{r2}}$ is actually optimizing the correlation. The difference in performance between SPoC$_{\text{r2}}$ and SPoC$_\lambda$ is strongest in the correlations between z and ϕ in high SNR regimes. In terms of pattern correlation and source reconstruction error the differences between SPoC$_{\text{r2}}$ and SPoC$_\lambda$ are negligible compared to the differences between the SPoC variants and the other methods.

Figure 4.4 compares activation patterns obtained from individual simulation runs under three different SNR conditions. The SNR levels shown here -6 dB, -12 dB, and -18 dB. Depicted are the activation pattern of the target source, and estimates thereof obtained from the different algorithms. This display gives a qualitative impression of the quantitative results shown in figure 4.3.

In real data analysis settings one cannot be sure if the target variable (z in this case) really corresponds to the activation of a component hidden in the data. This aspect is explored in simulations in which we varied the correlation between z and the true bandpower time-course of the target source. In this way we explored the robustness of the analysis approaches with

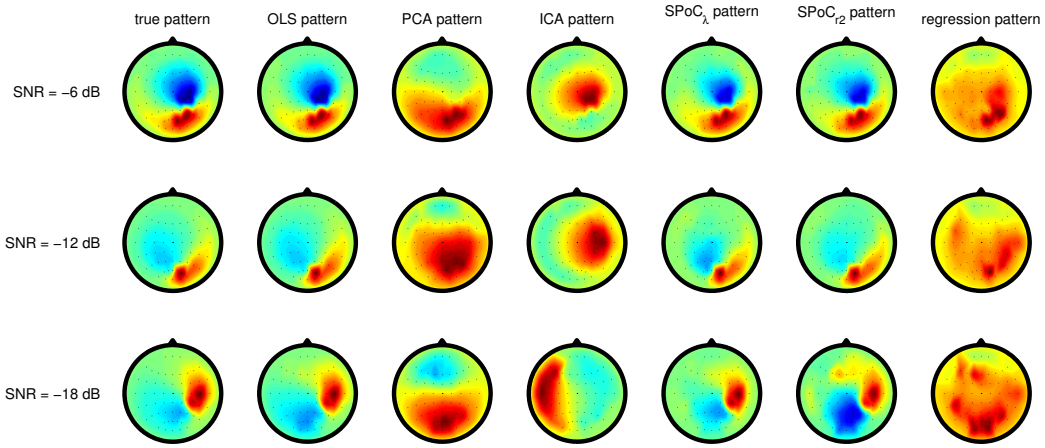


Figure 4.4: Simulation results. Spatial activation patterns from three separate simulation runs, each with a different signal-to-noise ratio (SNR). Each row correspond to one SNR, with the SNR level depicted furthest to the left. The left most column shows the activation pattern of the target source in the respective run. From left to right are the estimates thereof, as obtained from the different methods. Note that the OLS method requires knowledge of the actual time-course of the target source and is thus not applicable in practice.

respect to label noise. The results are depicted in figure 4.5. As expected, the performance degrades as the training signal shows less and less similarity with the activity of the target component. Note that the performance of the unsupervised methods (PCA and ICA) is also affected by this quantity, because the target component is selected from the set of PCA/ICA components based on bandpower correlation with the z variable on the training data.

The last aspect we investigate using simulations is the effect of the amount of training data that is available. In our simulations we use continuous data that is segmented into 500 ms long non-overlapping epochs. For all simulations discussed before, the amount of training epochs 240, which corresponds to 2 minutes of continuous data. Since trial-based analysis settings are common, we measure the amount of data used for training in epochs, rather than in time. The amount of training data required is explored by varying the number of training epochs, while keeping the number of test epochs constant (240). Figure 4.6 shows the results of these simulations. It can be observed that the SPoC algorithms surpass ICA once more than 100 to 150 training epochs (i.e. about one minute of data) are available. After that, it takes the SPoC algorithms about 250 epochs (about two minutes of data) more to reach their peak performance for this particular setting of simulation parameters.

4.2.2 Real data examples

The utility of the SPoC methods is further demonstrated using real EEG data. To this end, an experiment on steady-state auditory evoked potentials (SSAEPs) (Picton et al., 2003) was conducted with 11 participants, of which N=7 showed a SSAEP. The data from the remaining 4 participants was not used in this analysis.

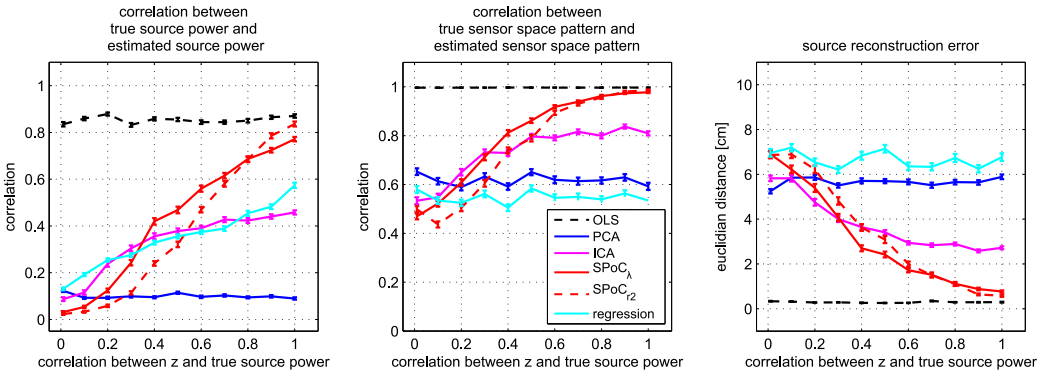


Figure 4.5: Simulation results. Here performance in terms of accuracy and interpretability is shown as a function of correlation between the target variable z and the bandpower time-course of the target source. The results document the performance of the methods in the presence of so-called label noise. The SNR was -10 dB in these simulations.

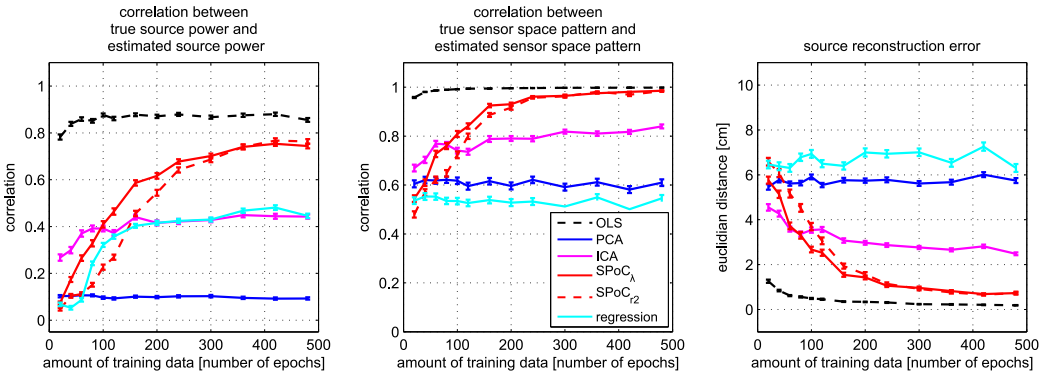


Figure 4.6: Simulation results. Here performance in terms of accuracy and interpretability is shown as a function of amount of training data available. Training data is measured in number of epochs, each of which was 500 ms long in these simulations. Epochs did not overlap. The SNR was set to -10 dB and the correlation between z and the bandpower of the target source were set to 0.7.

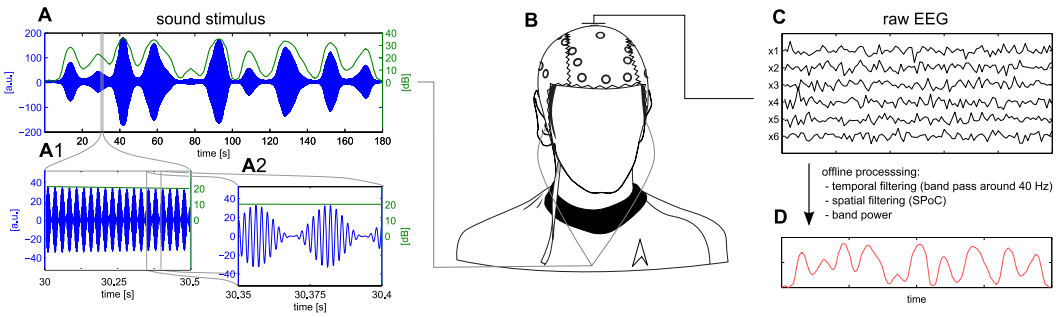


Figure 4.7: Auditory stimulus and experimental setup. (**A**) Three minute excerpt of the intensity modulated steady state stimulus. A 500 Hz sinusoid was multiplied with a 40 Hz raised cosine (see the 0.5 second and 0.05 second excerpts in **A**1 and **A**2, respectively). The slowly varying intensity modulation (green line in **A**, **A**1, and **A**2) was applied to the full length steady state stimulus. (**B**) Participants received the sound stimulus via in-ear headphones and EEG was measured concurrently. (**C**) The raw EEG was analyzed offline to extract an estimate of the intensity modulation (**D**).

Experimental Paradigm

The auditory stimulus consisted of a sinusoid with a carrier frequency of 500 Hz which was amplitude modulated with a 40 Hz raised cosine, thus resulting in the steady-state modulation. The resulting sound stimulus is referred to as the *steady-state stimulus*. It has been shown that such a stimulus induces a reliable steady state response in the auditory system, i.e. a significant increase in EEG/MEG power at the stimulation frequency (Galambos et al., 1981; Hari et al., 1989; Plourde et al., 1991; John et al., 2003; Picton et al., 2003). The SSAEP literature suggests a positive correlation between the amplitude of the evoked EEG response and the intensity of the steady-state stimulus when measured in decibel (dB) (Rodriguez et al., 1986; Plourde et al., 1991).

In our paradigm, we realized a continuous amplitude modulation of the sound stimulus by multiplying it with a slowly varying function, which we refer to as the *intensity modulation*. This function modulated the loudness of the stimulus between 10 and 35 dB relative to the subject- and ear specific hearing level (HL). The intensity modulation was created by low-pass filtering white noise with a cut-off frequency of 0.05 Hz, which yields a random, yet smoothly varying fluctuation. Before applying the intensity modulation to the sound stimulus, we equalized the histogram of the intensity modulation such that all sound intensity levels appear with equal probability. The beginning and the end of the sound stimulus was faded in/out to minimal intensity using a half cosine window of 10 seconds duration. Figure 4.7 illustrates the experimental setup, including the construction of the sound stimulus.

The experiment consisted of 3 blocks of 5 minutes continuous stimulation. Between each block, there was a short pause (less than 1 minute) for the participants to rest briefly. The sound stimulus was delivered using in-ear headphones and during the EEG recording participants were instructed to relax but keep their eyes open and to focus on the sound.

Data Acquisition

EEG signals were recorded using a Fast'n Easy Cap (EasyCap GmbH) with 63 wet Ag/AgCl electrodes placed at symmetrical positions based on the International 10–20 system. Channels were referenced to the nose. Electrooculogram (EOG) signals were recorded in addition but not used in the present context. Signals were amplified using two 32-channel amplifiers (Brain Products) and sampled at 1 kHz.

Data Analysis

For the offline analysis in MATLAB, the signals were low-pass filtered with a cutoff frequency of 90 Hz and subsequently down-sampled to 250 Hz. Additionally a notch filter around 50 Hz was applied to attenuate line noise. The down-sampled and notch filtered EEG data were then band pass filtered with a 3 Hz pass band centered on the steady-state frequency of 40 Hz, yielding a pass band from 39 to 41 Hz. The band pass filtered data was then segmented into consecutive epochs of 2 seconds length and 1 second overlap.

The SSAEP literature suggests a linear relationship between the stimulus intensity (measured in dB) and EEG amplitude at the stimulus frequency. Since the SPoC algorithms and linear regression work on power features (i.e. squared amplitude), we used the *squared* stimulus intensity as the target variable z.

PCA preprocessing (dimensionality reduction, retaining 99% of the variance) was employed for ICA as this improved the results compared to using ICA without PCA preprocessing.

In order the get an unbiased estimate of each of the methods' ability to model the stimulus intensity modulations, we employed a 10-fold chronological cross-validation procedure (Lemm et al., 2011). This means that the whole data was split up into 10 equally sized folds, of which 9 folds served as training data while the remaining fold was used for testing. This training/testing split was repeated such that each fold became the test fold once, yielding a correlation value for each split. Cross-validation was performed for all methods. The obtained correlations were transformed using Fisher's z-transform, averaged, and the mean z-value was then transformed back into a correlation value using the inverse of Fisher's z-transform.

Results

Fig. 4.8 depicts the results obtained from the auditory steady state experiment with a slowly changing intensity (i.e. loudness) modulation. Each colored line in the left part of the figure corresponds to a single participant, while the right part of the figure shows the average across participants. The SPoC algorithms and ICA return several components and therefore the plotted values refer to the cross-validated correlations obtained from the extracted component that had the highest correlation between its power time course and the sound intensity modulation. It can be seen that $\mathrm{SPoC}_{\mathrm{r2}}$ and SPoC_{λ} outperform ICA and regression in the majority of subjects. The SPoC algorithms yields statistically significant larger correlations between power time courses and the sound intensity modulation than

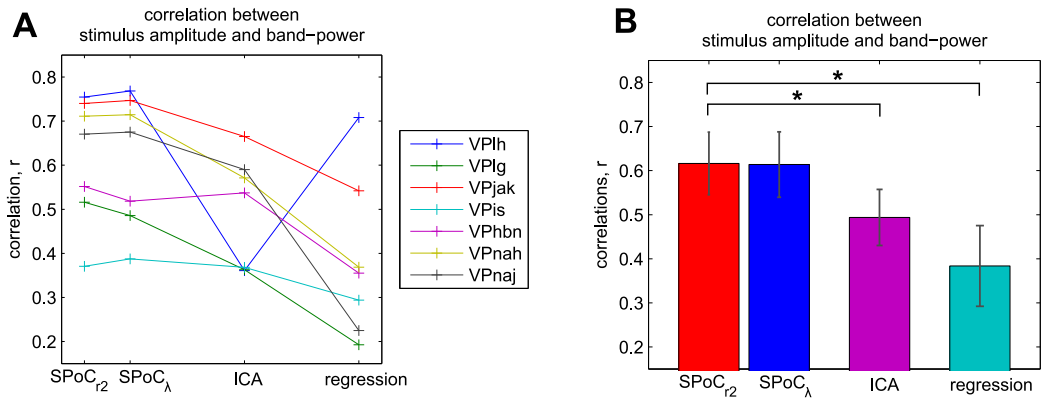


Figure 4.8: Real EEG data results: correlation between stimulus intensity and EEG power at the steady state frequency. (**A**) Each colored line corresponds to a participant and shows the obtained correlations for all methods (mean over cross-validation folds). (**B**) Same information as in **A**, averaged over subjects. Error bars depict the standard error of the mean and a black star indicates statistically significant difference at $p<0.05$.

ICA or regression ($p < 0.05$, Wilcoxon rank sum test). Furthermore, on this data set the performance of SPoC$_\lambda$ is statistically indistinguishable from the performance of SPoC$_{r2}$.

For the two SPoC algorithms, the best correlating component was always the first component with respect to the ordering imposed by the respective objective function. A permutation analysis[2] revealed that for the data of all but one participant the power time course of the respective first SPoC component captured the variation due to the target function.

Fig. 4.9 shows the spatial patterns corresponding to the best spatial filters for each subject. Best spatial filter here means the spatial filter $\mathbf{w}$ that yielded the largest correlation between the power time course of $\mathbf{w}^\top\mathbf{x}$ (i.e. the power time course of the filtered signal) and z. Please note that the polarity of the spatial patterns (as well as of the corresponding filters) is arbitrary. For each pattern, the polarity was set such that the pattern value at EEG electrode Cz is positive.

Fig. 4.10 shows more detailed results for a representative participant (VPnaj). These plots show channel-wise correlations plotted as a scalp map; the spatial patterns of highest correlating SPoC$_{r2}$, SPoC$_\lambda$, and ICA components; as well as the power spectra of a single EEG channel (Fz) and the spectra obtained after spatial filtering with the corresponding SPoC$_{r2}$, SPoC$_\lambda$, or ICA filter. Please note that for all subjects the SPoC$_\lambda$ filter that maximized the covariance (i.e. the objective function of SPoC$_\lambda$) also exhibited the maximal correlation between the power time course and stimulus intensity. It can be seen that the channel-wise correlations are low in magnitude and that the pattern of correlation values shows little resemblance with the components obtained from the spatial filtering methods. Between SPoC$_{r2}$, SPoC$_\lambda$, and ICA, the obtained patterns are quite similar, indicating that

[2]The analysis was re-run 500 times, each time using a newly shuffled version of the target function. This resulted in a distribution of correlation/covariance values under the null-hypothesis that the target function and the EEG power do not share the same time course.

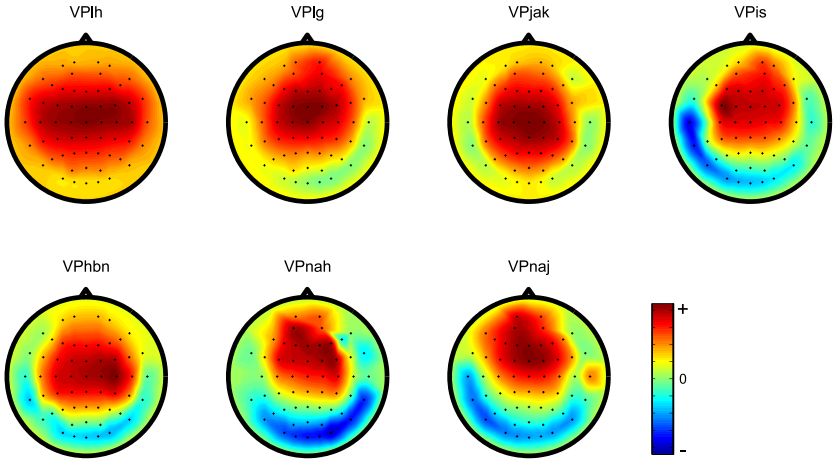


Figure 4.9: Real EEG data results: spatial patterns obtained by $SPoC_{r2}$ for each participant. The spatial patterns correspond to the filters that maximize the correlation between power time course and stimulus intensity.

the same source (or set of sources) has been extracted by the algorithms. The second row of plots in Fig. 4.10 shows the offset-aligned power spectra of EEG channel Cz and the respective best $SPoC_{r2}$, $SPoC_{\lambda}$, and ICA components (corresponding to the spatial patterns above). The spatial filtering methods show a much clearer peak at the steady-state frequency compared to the individual recording channel. The peak is most pronounced in the component extracted by $SPoC_{r2}$, thus yielding the highest signal-to-noise ratio for this particular participant.

4.3 Discussion

In this chapter we presented a novel multivariate approach for the extraction of oscillatory sources showing a co-modulation of their power with the target function, the latter being for instance reaction time, hit rate or some physical properties of the sensory stimuli (e.g. intensity). The SPoC approach is the first to explicitly address the problem of component extraction for band power correlation/covariance. Using two implementations of our approach (the $SPoC_{\lambda}$ and the $SPoC_{r2}$ algorithm) we were able to show that it performed better than other methods commonly used for the investigation of a relationship between behavioral measures and a power of oscillations (sensor space regressions, ICA).

SPoC showed promising results in the extraction of auditory sources generating steady-state responses. The extracted patterns were consistent with ERP-type analysis of auditory steady state responses in the 40 Hz range. Herdman et al. (2002) investigated the source activity underlying the responses to 39 Hz modulated tones. They found a vertically oriented dipolar pattern with maxima (minima) at mid-frontal electrode positions and corresponding minima (maxima) at posterior neck positions, where the sign of minima and maxima depends on

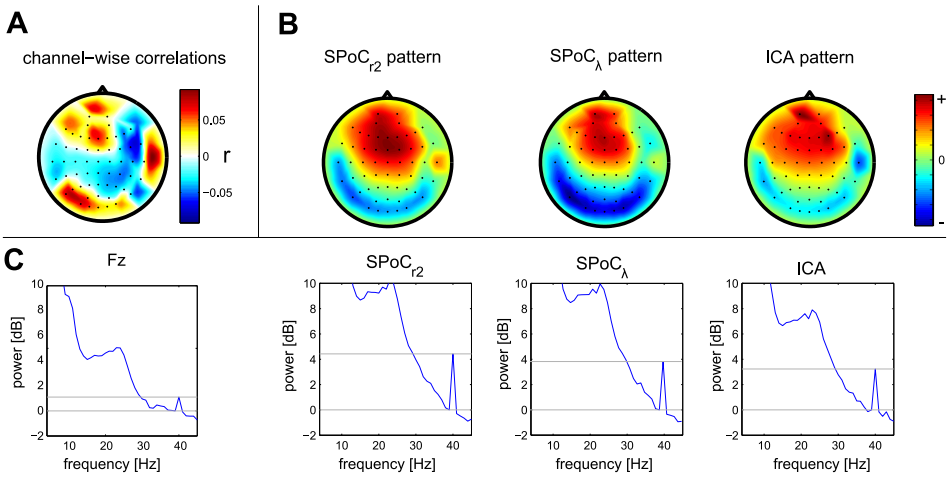


Figure 4.10: Real EEG data results: spatial patterns and power spectra for a representative participant (VPnaj). (**A**) Channel-wise correlation between band power and sound intensity. (**B**) Spatial patterns of the best SPoC$_{r2}$, SPoC$_{\lambda}$, and ICA component. (**C**) Offset-aligned Power spectra of a single EEG channel (Fz), as well as of the components depicted in the top row. Grey lines in the power spectra plot indicate the difference between power in the 39 Hz and 40 Hz bin, i.e. the signal-to-noise ratio in the spectral domain.

the phase of the stimulus. The obtained spatial SPoC patterns corresponded well with the findings of Herdman et al. (2002).

Comparing the results obtained on the real EEG data to those obtained on the simulated data, we find that the performance differences between SPoC and ICA on the real data are comparable to those found in simulations with moderately high SNR. This indicates that the sources activated in the auditory steady state paradigm are relatively strong compared to background activity in the frequency band of interest. However, the significantly larger correlations obtained with SPoC underline its ability to extract the signal of interest while effectively suppressing unrelated noise sources at the same time. This makes SPoC a valuable tool in scenarios in which the time course of the target variable is to be predicted from ongoing EEG in an online manner.

Below we elaborate on some technical aspects of the SPoC algorithms. If an analysis setting requires the exploration of one or more parameters and thereby multiple runs of the analysis method, then processing speed of the method might be an issue. Some BSS methods require a number of iterations which may take minutes to converge on a standard computer. For a single run only, the range of minutes should not, however, be a problem. However, if the number of runs increases greatly (e.g. for bootstrapping (Meinecke et al., 2002)) minutes can easily become many hours on single processors. Thus it is important to point out that the generalized eigenvalue computation in the SPoC$_{\lambda}$ algorithm takes only fractions of a second, and that is why SPoC$_{\lambda}$ unproblematically allows for (a) extensive parameter studies, or (b) bootstrapping efforts or (c) online evaluations within a time window with only moderate processing time demands.

As a technical side remark we would like to also mention that the SPoC$_\lambda$ algorithm is intimately related to the Common Spatial Pattern (CSP) algorithm family (Blankertz et al., 2008). CSP is the current most popular method in Brain-Computer Interface systems, which are based on oscillatory brain signals (Lemm et al., 2011; Blankertz et al., 2007). When the target variable z is binary, classical CSP is obtained as a special case of SPoC$_\lambda$. One may thus view SPoC$_\lambda$ as a regression extension of CSP to continuous target variables. In such a regression scenario, CSP could still be used but it would require a form of binning of the target variable (e.g. mean- or median split, or using percentiles), which might be arbitrary. Both SPoC algorithms, however, are specifically designed for continuous target variables and therefore do not require such preprocessing.

The SPoC approach is intended to be used on multichannel EEG and MEG data. However, it is not limited to the application of non-invasive imaging methods and should perform equally well for invasive recordings. A specific application scenario for SPoC would be studies with intra-cortical electrodes in epilepsy patients and recordings obtained from deep brain structures such as Globus Pallidus and Subthalamic Nucleus in patients with Parkinson's Disease. In fact, given a special clinical interest in understanding brain mechanisms of the neurological disorders, a precise localization of pathological neuronal networks would be a great advantage, such as for instance in clarifying the neuronal generators of tremor in patients with Parkinson's Disease (Wichmann and Delong, 2011). In this case a tremor can be used as a target variable in order to extract corresponding sources of beta or high-frequency oscillations in the thalamo-cortical-basal network.

Note that SPoC properly implements the commonly accepted generative model of EEG/MEG and therefore it is possible to meaningfully interpret its results within this generative model. This also allows subsequent source localization (Baillet et al., 2001; Haufe et al., 2008, 2011) or further multimodal processing (Fazli et al., 2012; Bießmann et al., 2011).

In summary, SPoC is an approach that enables a reliable and fast extraction of neuronal oscillations, whose power time course co-modulates with an external target function. Because of SPoC's superiority to other standard techniques, we advocate its use for recovering associations between cognitive/motor variables and neuronal activity.

Chapter 5

Multimodal Source Power Co-modulation

In this chapter we extend the Source Power Co-modulation idea to the multivariate and, more importantly, to the *multimodal* case. To this end, we assume to be confronted with a multivariate oscillatory dataset $\mathbf{x}$ as well as another multivariate dataset, denoted by $\mathbf{y}$. The task is now to extract source components from both datasets that are functionally connected to each other. Specifically, we seek to uncover component pairs ($s_{\mathbf{x}}$ and $s_{\mathbf{y}}$) such that the bandpower dynamics of $s_{\mathbf{x}}$, denoted by $\phi_{\mathbf{x}}$, co-modulate with the time-course of a corresponding $s_{\mathbf{y}}$.

Part of this task is solved using ideas that we developed in the previous chapter, especially concerning the extraction of components from $\mathbf{x}$. However, instead of having a univariate target function z available, we must now extract this signal from the multivariate dataset $\mathbf{y}$. In all of the following considerations we assume that both $\mathbf{x}$ and $\mathbf{y}$ can be modeled using the linear generative model described in chapter 2.2.1, with modality specific mixing coefficients $\mathbf{A}_{\mathbf{x}}$ and $\mathbf{A}_{\mathbf{y}}$, as well as source time-courses $\mathbf{s}_{\mathbf{x}}$ and $\mathbf{s}_{\mathbf{y}}$. Thus, the approach we propose in this chapter respects the linear generative model that was outlined in chapter 2.2. Thereby, our approach allows for meaningful interpretation of the resulting components, not only with respect to functional aspects but also with respect to anatomical locations.

The method we develop in this chapter assumes a set of underlying sources whose activity is visible in both imaging modalities. The estimation of source components is based on maximization of a co-modulation term and therefore we refer to our method as *multimodal source power co-modulation* (mSPoC). The objective function we define represents our novel approach to the problem. The way we actually optimize the objective, however, is rather generic and can in principle replaced by a different optimization procedure.

A conceptual comparison of (i) blind source separation methods applied in the multimodal context, (ii) the previously introduced method canonical correlation analysis (CCA), and (iii) our novel method mSPoC is provided in Fig. 5.1.

5.1 The mSPoC algorithm

5.1.1 Defining the objective function

Similar to our approach in the previous chapter, we employ a backward modeling, i.e. spatial filtering, approach to extract the component time-courses of interest from the multivariate

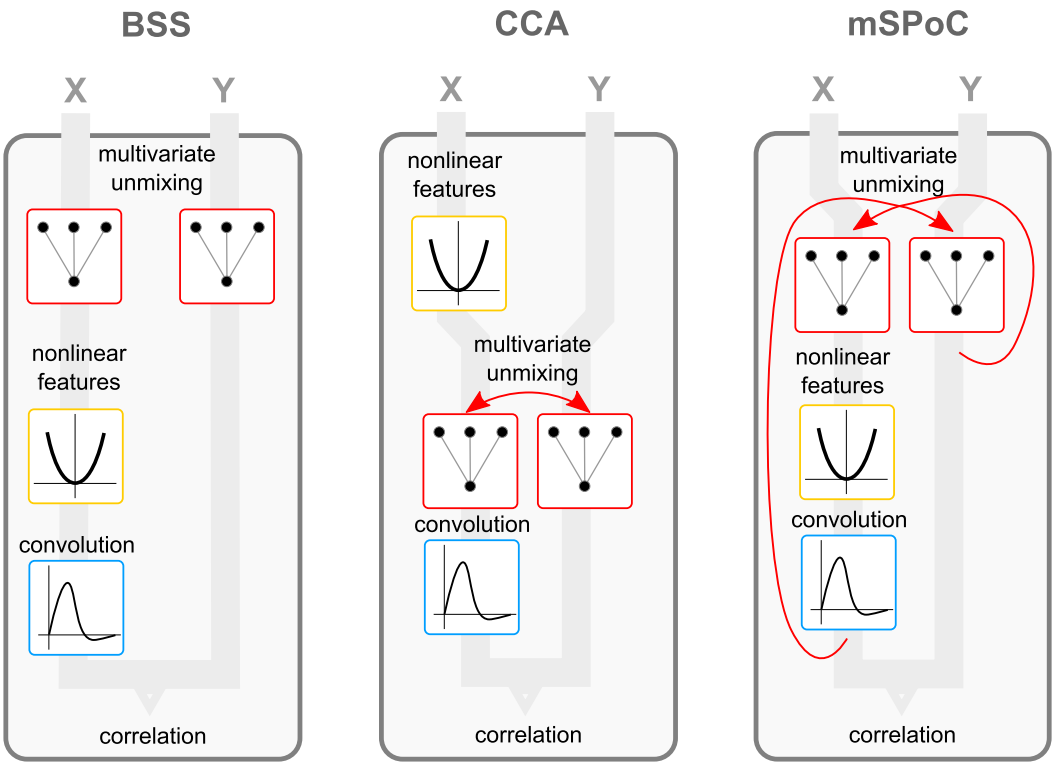


Figure 5.1: Illustration of different approaches to symmetrical fusion of EEG spectral power and multivariate hemodynamics. The input to all three approaches is the multivariate and bandpass filtered EEG/MEG data $\mathbf{x}$ as well as a multivariate time series of, for example, NIRS of fMRI measurements. Processing steps are organized from top to bottom. **Left:** An approach that is based on blind source separation (BSS) methods that are applied to each modality independently. This approach is in line with the generative model we assume for both modalities. However, the modality-specific BSS methods do not inform each other during the decomposition and thus potentially miss out on available information. Furthermore, after the decomposition step, a matching of components across modalities is required as well as an additional estimation step for the hemodynamic response function. **Middle:** An approach based on canonical correlation analysis (CCA) or partial least squares (PLS). In this approach, the decomposition into components is informed by the respective other modality and an additional matching of components across modalities is not required. The crucial drawback of this approach is that on the EEG/MEG side, bandpower has to be computed on channel-level, which leads to a number of shortcomings that have been outlined in chapter 3. **Right:** Our proposed approach called multimodal source power co-modulation (mSPoC) that combines advantages of the two other approaches (conformity with the generative model as well as maximal use of cross-modal information) while avoiding their shortcomings.

datasets. We denote the extracted signals by

$$\hat{s}_{\mathbf{x}} = \mathbf{w}_{\mathbf{x}}^{\top}\mathbf{x} \text{ and} \tag{5.1}$$
$$\hat{s}_{\mathbf{y}} = \mathbf{w}_{\mathbf{y}}^{\top}\mathbf{y}\,, \tag{5.2}$$

where the variables $\hat{s}_{\mathbf{x}}$ and $\hat{s}_{\mathbf{y}}$, respectively, denote the component time-series extracted from the data by means of spatial filters $\mathbf{w}_{\mathbf{x}}$ and $\mathbf{w}_{\mathbf{y}}$.

We assume $\mathbf{x}$ and $\mathbf{y}$ to be mean-free and temporally aligned. We further assume $\mathbf{x}$ to be bandpass filtered for a frequency band of interest and that it can be segmented into small time windows, called epochs, with index e for $e \in \{1, ..., N_e\}$. The individual epochs should be long enough to allow for a reasonable estimate of spectral power of $\mathbf{x}$, yet short enough so that $\mathbf{y}$ changes only very little within each individual epoch[1]. Finally, $\mathbf{y}$ should be sub-sampled to epoch time resolution such that it can be indexed with the epoch index e, i.e. we have $\mathbf{y} = \{\mathbf{y}(1), ..., \mathbf{y}(e), ..., \mathbf{y}(N_e)\}$.

The epoch-wise power of $\hat{s}_{\mathbf{x}}$ is approximated by the variance within each epoch, which leads to:

$$\begin{aligned}\phi_{\mathbf{x}}(e) &= \mathrm{Var}\left[\hat{s}_{\mathbf{x}}(t)\right](e) \\ &= \mathbf{w}_{\mathbf{x}}^{\top}\mathbf{C}_{\mathbf{xx}}(e)\mathbf{w}_{\mathbf{x}},\end{aligned} \tag{5.3}$$

where $\mathbf{C}_{\mathbf{xx}}(e)$ denotes the covariance matrix of $\mathbf{x}$ in the epoch e.

We model the possibly time-delayed coupling between $\phi_{\mathbf{x}}$ and $\hat{s}_{\mathbf{y}}$ using a finite impulse response (FIR) filter approach. The FIR filter coefficients are stored in temporal weight vector $\mathbf{w}_\tau$ that is applied to $\phi_{\mathbf{x}}$. Thus we define the temporally filtered power time-series

$$\hat{h}(\phi_{\mathbf{x}})(e) = \sum_{i=1}^{N_\tau} w_{\tau i}\, \phi_{\mathbf{x}}(e - \tau_i)\,, \tag{5.4}$$

where the scalar variable $w_{\tau i}$ denotes the i-th element of the vector $\mathbf{w}_\tau = (w_{\tau 1}, ..., w_{\tau N_\tau})^{\top}$, τ_i for $i \in \{1, ..., N_\tau\}$ is a set of suitable pre-chosen time-lags, and N_τ is the total number of time-lags to be taken into account.

The last equation depends both on $\mathbf{w}_\tau$ and $\mathbf{w}_{\mathbf{x}}$. In order to simplify the optimization (see next section), we give two alternative but equivalent expressions for $\hat{h}(\phi_{\mathbf{x}})$. The first one is based on the observation that the spatial filter $\mathbf{w}_{\mathbf{x}}$ does not depend on the index i in the last line of the Eq. (5.4) and, thus, can be moved out of the sum. Doing so leads to

$$\begin{aligned}\hat{h}(\phi_{\mathbf{x}})(e) &= \sum_{i=1}^{N_\tau} w_{\tau i}\; \mathbf{w}_{\mathbf{x}}^{\top}\mathbf{C}_{\mathbf{xx}}(e - \tau_i)\mathbf{w}_{\mathbf{x}} \\ &= \mathbf{w}_{\mathbf{x}}^{\top}\left(\sum_{i=1}^{N_\tau} w_{\tau i}\, \mathbf{C}_{\mathbf{xx}}(e - \tau_i)\right)\mathbf{w}_{\mathbf{x}} \\ &= \mathbf{w}_{\mathbf{x}}^{\top}\; \hat{h}(\mathbf{C}_{\mathbf{xx}})(e)\; \mathbf{w}_{\mathbf{x}}\,,\end{aligned} \tag{5.5}$$

[1] In most practical scenarios this setting is very natural, especially if $\mathbf{x}$ and $\mathbf{y}$ are acquired with different sampling rates. A typical example would be concurrent EEG and fNIRS, where half a second represents an appropriate epoch length.

where $\hat{h}(\mathbf{C_{xx}})(e)$ represents the element-wise convolution of the covariance time-series $\mathbf{C_{xx}}(e)$ with the FIR filter $\mathbf{w}_\tau$. Please note the resemblance between Eq. (5.5) and Eq. (5.3). The second alternative version of Eq. (5.4) puts more emphasis on $\mathbf{w}_\tau$ and reads

$$\hat{h}(\phi_\mathbf{x})(e) \quad = \quad \mathbf{w}_\tau^\top \Phi_\mathbf{x} \,, \tag{5.6}$$

where $\Phi_\mathbf{x} \in \mathbb{R}^{N_\tau \times (N_e - N_\tau)}$ is the temporally embedded version of $\phi_\mathbf{x}$:

$$\Phi_\mathbf{x} \quad = \quad \begin{bmatrix} \mathrm{sh}(\phi_\mathbf{x}, \tau_1) \\ \vdots \\ \mathrm{sh}(\phi_\mathbf{x}, \tau_{N_\tau}) \end{bmatrix} , \tag{5.7}$$

where we define $\mathrm{sh}(\phi_\mathbf{x}, \tau_i)$ to be a copy of $\phi_\mathbf{x}$ with all elements shifted by the time-lag τ_i, i.e. $\mathrm{sh}(\phi_\mathbf{x}, \tau_i)(e) = \phi_\mathbf{x}(e - \tau_i)$.

Equipped with these definitions, we now formulate an mSPoC objective function using covariance as the co-modulation function:

$$\max_{\mathbf{w_x}, \mathbf{w_y}, \mathbf{w}_\tau} \mathrm{Cov}\left(\mathbf{w}_\tau^\top \Phi_\mathbf{x}, \mathbf{w}_\mathbf{y}^\top \mathbf{y}\right), \tag{5.8}$$

subject to the following norm constraints

$$\mathbf{w}_\mathbf{x}^\top \left\langle \mathbf{x}\mathbf{x}^\top \right\rangle \mathbf{w}_\mathbf{x} \quad = \quad \mathbf{w}_\mathbf{x}^\top \mathbf{C_{xx}} \mathbf{w_x} = 1, \tag{5.9}$$

$$\mathbf{w}_\mathbf{y}^\top \left\langle \mathbf{y}\mathbf{y}^\top \right\rangle \mathbf{w}_\mathbf{y} \quad = \quad \mathbf{w}_\mathbf{y}^\top \mathbf{C_{yy}} \mathbf{w_y} = 1, \tag{5.10}$$

$$\mathbf{w}_\tau^\top \mathbf{w}_\tau \quad = \quad 1, \tag{5.11}$$

In Dähne et al. (2013) we have shown that the objective function can be further expanded and then expressed in terms of a deflation of a four dimensional covariance tensor $\mathbf{C_{xxy}}_\tau \in \mathbb{R}^{N_x \times N_x \times N_y \times N_\tau}$. While this leads to a conveniently concise expression of the objective function, the actual computation of the covariance tensor is less convenient. In fact, the size of this object, i.e. the total number of elements in $\mathbf{C_{xxy}}_\tau$, scales with $\mathcal{O}(N_x \cdot N_x \cdot N_y \cdot N_\tau)$, which is not particularly desirable. However, as we show in the next section, the covariance tensor does not have to computed explicitly and therefor we do not make any use of it in further derivations. In fact, it turns out that we can break the objective function down into connected sub-problems that can each be solved much more efficiently, leading to space requirements in the range of $\mathcal{O}(N_x^2 + N_y^2 + N_\tau^2)$ only.

5.1.2 Optimization of the objective function

In order to optimize the mSPoC objective function, we follow a typical analytical optimization approach. That is, we first define the Lagrangian of the original objective by including the constraints:

$$\begin{aligned} L(\mathbf{w_x}, \mathbf{w_y}, \mathbf{w}_\tau, \lambda_\mathbf{x}, \lambda_\mathbf{y}, \lambda_\tau) \quad = \quad & \mathrm{Cov}\left(\mathbf{w}_\tau^\top \Phi_\mathbf{x}, \mathbf{w}_\mathbf{y}^\top \mathbf{y}\right) \\ & + \lambda_\mathbf{x} \left(1 - \mathbf{w}_\mathbf{x}^\top \mathbf{C_{xx}} \mathbf{w_x}\right) \\ & + \lambda_\mathbf{y} \left(1 - \mathbf{w}_\mathbf{y}^\top \mathbf{C_{yy}} \mathbf{w_y}\right) \\ & + \lambda_\tau \left(1 - \mathbf{w}_\tau^\top \mathbf{w}_\tau\right) \,, \end{aligned} \tag{5.12}$$

where $\lambda_{\mathbf{x}}$, $\lambda_{\mathbf{y}}$, and λ_τ are unknown scalar variables.

Then we compute the partial derivatives with respect to $\mathbf{w_x}$, $\mathbf{w_y}$, and $\mathbf{w_y}$, respectively, set those equal to zero and then solve for $\mathbf{w_x}$, $\mathbf{w_y}$, and $\mathbf{w_y}$. This approach leads to the following set of coupled equations:

$$\left\langle \hat{h}(\mathbf{C_{xx}})(e)\, \hat{s}_{\mathbf{y}}(e) \right\rangle \mathbf{w_x} = \lambda_{\mathbf{x}} \mathbf{C_{xx}} \mathbf{w_x} \tag{5.13}$$

$$\frac{1}{2} \mathbf{C_{yy}}^{-1} \left\langle \hat{h}(\phi_{\mathbf{x}})(e)\, \mathbf{y}(e) \right\rangle = \lambda_{\mathbf{y}} \mathbf{w_y} \tag{5.14}$$

$$\frac{1}{2} \left\langle \Phi_{\mathbf{x}}(e)\, \hat{s}_{\mathbf{y}}(e) \right\rangle = \lambda_\tau \mathbf{w}_\tau \,, \tag{5.15}$$

in which $\langle \cdot \rangle$ denotes the average taken over epochs. These equations hold simultaneously in local optima of the objective function.

Note that $\left\langle \hat{h}(\mathbf{C_{xx}})(e)\, \hat{s}_{\mathbf{y}}(e) \right\rangle$ in Eq. (5.13) evaluates to a matrix of size $N_x \times N_x$ that depends on $\mathbf{w_y}$ and $\mathbf{w}_\tau$. This in turn means that for a given $\mathbf{w_y}$ and $\mathbf{w}_\tau$, the optimal $\mathbf{w_x}$ is found as the solution to a generalized eigenvalue problem. In fact, Eq. (5.13) is the solution to the SPoC$_\lambda$ objective function that we have seen in the previous chapter, with $\hat{s}_{\mathbf{y}}$ now in the role of z and the filtered covariance time-series $\hat{h}(\mathbf{C_{xx}})(e)$ in the role of $\mathbf{C_{xx}}(e)$. Furthermore, for a given $\mathbf{w_x}$ and $\mathbf{w}_\tau$, we see that the objective function becomes a linear regression with the univariate $\hat{h}(\phi_{\mathbf{x}})$ as the dependent variable and $\mathbf{y}$ as the multivariate regressor, or predictor variables. Eq. (5.14) gives us an expression for $\mathbf{w_y}$ as the ordinary least squares solution to the regression problem. Finally, a similarly simple solution is obtained for the optimal $\mathbf{w}_\tau$, given fixed $\mathbf{w_x}$ and $\mathbf{w_y}$.

The dependency structure between $\mathbf{w_x}$, $\mathbf{w_y}$, and $\mathbf{w}_\tau$ can be used to form a simple optimization scheme, in which one would start with a random initialization of all weight vectors and then iterate Eq. (5.13), Eq. (5.14), and Eq. (5.15) until convergence. In our optimization procedure, we simplify the structure even further by noting that for given $\mathbf{w_x}$, estimates of both $\mathbf{w_y}$ and $\mathbf{w}_\tau$ can be obtained simultaneously using CCA, which in turn can be used to update $\mathbf{w_x}$ using SPoC$_\lambda$. Thus, in our implementation of mSPoC we make use of the previously discussed methods CCA and SPoC$_\lambda$, applying them iteratively to update $\mathbf{w_x}$, $\mathbf{w_y}$, and $\mathbf{w}_\tau$ until a suitable convergence criteria is met. Criteria for convergence can be the difference between consecutive objective function values or a fixed number of iterations for example. In our implementation we use a mixture of both, i.e. we compute the correlation between the current estimates of $\hat{h}(\phi_{\mathbf{x}})$ and $\hat{s}_{\mathbf{y}}$ after each iteration and stop if either the difference is smaller than some predefined threshold or a maximum number of iterations is reached.

Pseudo-code for the core part of our implementation, i.e. the extraction of $K = 1$ component from $\mathbf{x}$ and $\mathbf{y}$ is shown in Algorithm 1. In order to extract more than one component pair from the data, the case $K > 1$, we employ a deflation scheme that is outlined in chapter A.3.

Algorithm 1 Pseudocode of the mSPoC algorithm for $K = 1$ component pair. More component pairs are obtained using a deflation approach, i.e. by projecting out previously obtained components and repeating the method with deflated data.

Require: $\mathbf{x}$ band-pass filtered and $\mathbf{y}$ centered (optional: dimensionality reduction using SSD for $\mathbf{x}$ and (kernel-) PCA for $\mathbf{y}$)

function mSPoC($\mathbf{x}$, $\mathbf{y}$)
  $C_{xxe} \leftarrow$ compute covariance matrix time-series from $\mathbf{x}$
  for n restarts **do**
    $\mathbf{w_x} \leftarrow$ random initialization
    $\phi_\mathbf{x} \leftarrow$ power time-course from $\mathbf{w_x}$ and C_{xxe}
    while not converged **do**
      $\Phi_\mathbf{x} \leftarrow$ temporal embedding of $\phi_\mathbf{x}$
      $\mathbf{w}_\tau, \mathbf{w_y} \leftarrow \mathrm{CCA}(\Phi_\mathbf{x}, \mathbf{y})$
      $\hat{h}(C_{xxe}) \leftarrow$ filter C_{xxe} with $\mathbf{w}_\tau$
      $\hat{s}_\mathbf{y} \leftarrow \mathbf{w_y}^\top \mathbf{y}$
      $\mathbf{w_x} \leftarrow \mathrm{SPoC}_\lambda(\hat{h}(C_{xxe}), s_\mathbf{y})$
      $\phi_\mathbf{x} \leftarrow$ power time-course from $\mathbf{w_x}$ and C_{xxe}
      check convergence criteria
    end while
    store $\mathbf{w_x}$, $\mathbf{w_y}$, and $\mathbf{w}_\tau$ in a list along with correlation between $\hat{h}(\phi_\mathbf{x})$ and $\hat{s}_\mathbf{y}$
  end for
  return $\mathbf{w_x}$, $\mathbf{w_y}$, and $\mathbf{w}_\tau$ that yielded highest correlation
end function

5.2 Validation

5.2.1 Simulations

The data generation processes and the evaluation metrics are identical to what was described in chapter 3.2. The methods are compared in terms of *accuracy* and *interpretability*, where accuracy describes how well the functional activation of the target source components is extracted and interpretability reflects the similarity of the estimated activation patterns to those of the true target components.

For each simulation run, we created the datasets $\mathbf{x}$ and $\mathbf{y}$ according to the generative model given in Eq. (2.1), where the details of the data generation have been described in chapter 3.2. The number of simulated channels was $N_x = N_y = 25$, and a total of $K_x = K_y = 31$ sources was simulated for each data set. One source from $\mathbf{x}$ was functionally coupled with one source from $\mathbf{y}$, such that the envelope of this *target source* in $\mathbf{x}$ reflected the time-source of the respective other *target source* in $\mathbf{y}$. The time courses (and envelopes) of all other sources were completely independent to each other and to the target sources. These independent sources are referred to as background sources.

While the $\mathbf{x}$ data consisted of band-limited oscillations (8 Hz to 12 Hz), the $\mathbf{y}$ signal was

simulated to vary on slower time-scale in order to allow for correlations between amplitude modulations of the target $\mathbf{x}$ component and the corresponding target $\mathbf{y}$ component. To this end, all $\mathbf{y}$ source component signals are in fact amplitude modulations that have been obtained from band-limited oscillations, generated by the same function that generated the source components of the $\mathbf{x}$ signals. No time delay was introduced between the source components from the two simulated modalities.

Consistent with the simulations of the preceding chapters, we assessed the performance of the compared methods with respect to (i) varying signal-to-noise ratios (SNR) between the target source and the background sources, (ii) varying amounts of training data, and (iii) to the true degree of coupling between the target source components from $\mathbf{x}$ and $\mathbf{y}$, i.e. to the true correlation between $\mathbf{x}$ bandpower and $\mathbf{y}$ time-course. This last aspect assesses how the algorithms perform when the underlying assumption (functional coupling between modalities) are no longer valid.

The amount of training and test data was five and two minutes, respectively, for each of the simulation runs in which either the SNR or the true underlying coupling was varied. The $\mathbf{x}$ data was segmented into non-overlapping epochs of 500 ms length, while the $\mathbf{y}$ data was sub-sampled such that for each epoch in $\mathbf{x}$ there was a corresponding time sample in $\mathbf{y}$. For the simulation runs in which the amount of training was varied, the SNR was set to -10 dB. For each parameter setting the simulations were repeated 150 times with newly generated data.

Data analysis

We benchmarked mSPoC against CCA, ICA, and ordinary-least-squares (OLS).

OLS was applied to each modality separately and, being a supervized method, it received the time-courses of the true target sources as additional training data. Thus, OLS again serves as an empirical upper bound in the performance comparison.

For CCA, the $\mathbf{x}$ data was non-linearly preprocessed by computing the channel-wise variance within epochs. This resulted in channel-wise variance time-courses of $\mathbf{x}$ that matched the sampling frequency of the $\mathbf{y}$ data, which in turn allowed for application of standard CCA. The first CCA component pair, which is the pair that shows highest correlation on training data, was then used to asses its performance on the test data.

ICA was applied to the training data of each modality separately. Thereafter, the bandpower modulations of the ICA components obtained on the $\mathbf{x}$ training data were computed and correlated with the time-courses of the $\mathbf{y}$ data ICA components. The pair that yielded highest correlation on the training data was then assessed on the test data.

Results

Simulation results are depicted in figure 5.2, figure 5.3, and figure 5.4.

Figure 5.2 shows the performance of the compared methods as a function of SNR for two different settings of true underlying coupling (correlation 1 and correlation 0.7). mSPoC is

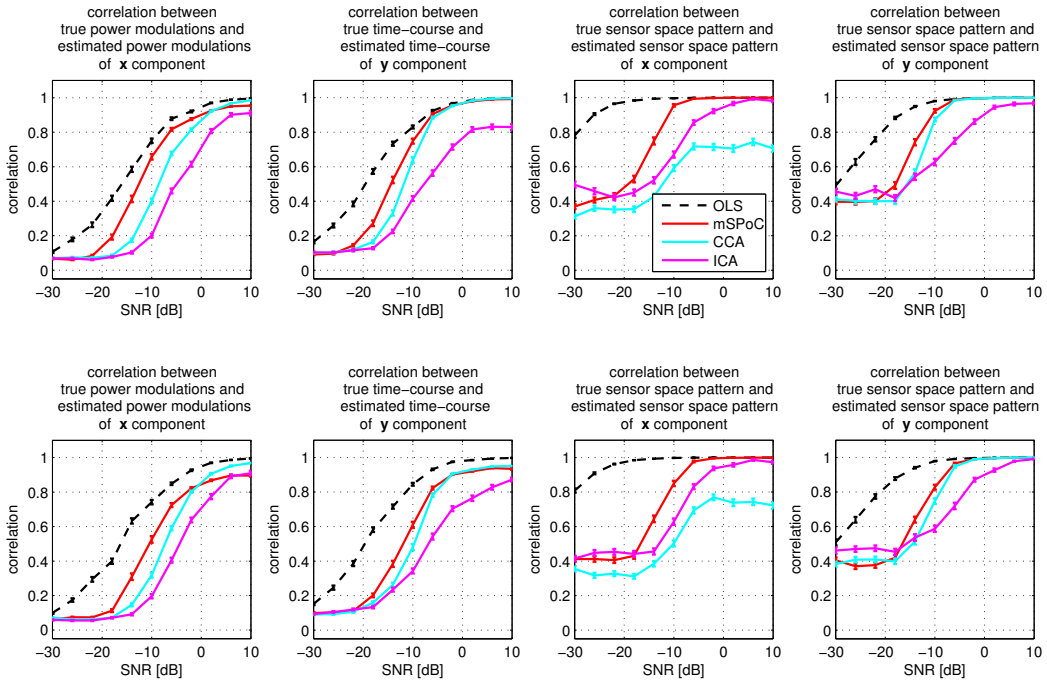


Figure 5.2: Simulation results. Here performance in terms of accuracy (first and second column) and interpretability (third and fourth column) is shown as a function of signal-to-noise ratio. In the top row of the plots, the true correlation between the bandpower modulations of the target **x** component and the time-course of the true y component was 1, while in the simulations depicted in the bottom row, the true correlation was 0.7.

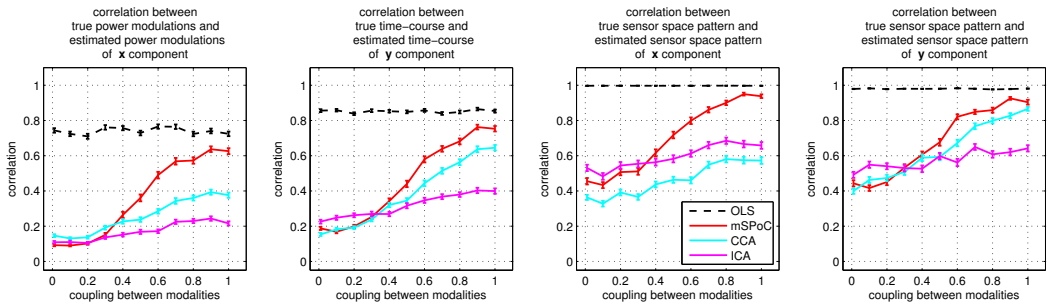


Figure 5.3: Simulation results. Here performance in terms of accuracy and interpretability is shown as a function of true underlying coupling, i.e. the correlation between the bandpower modulations of the target **x** component and the time-course of the target **y** component. The SNR was -10 dB in these simulations.

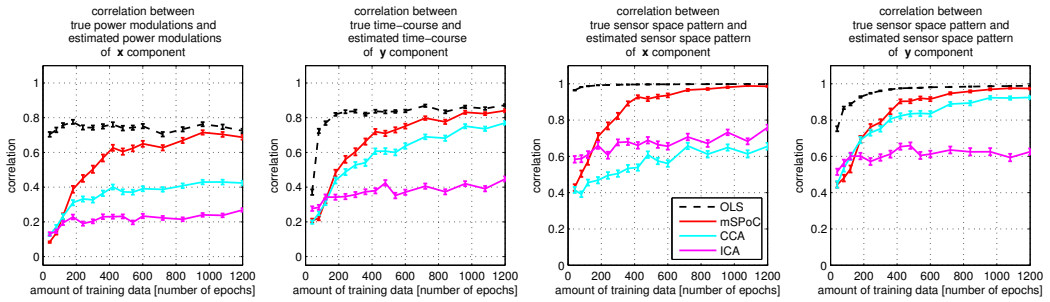


Figure 5.4: Simulation results. Here performance in terms of accuracy (first and second column) and interpretability (third and fourth column) is shown as a function of amount of training data available. Training data is measured in number of epochs of the **x** modality, each of which was 500 ms long and corresponded to the number of samples in the **y** modality. Epochs did not overlap and the SNR was set to -10 dB.

able to outperform CCA and ICA over wide ranges of SNR, particularly when the SNR is low. The performance differences between mSPoC and CCA are most clearly visible for the results obtained on the **x** data, because this part of the multimodal dataset is not accurately modeled by CCA. This inadequacy of CCA becomes most obvious in the interpretability metric (similarity between true and estimated activation patterns), where it is outperformed by both mSPoC and ICA.

The effect of varying degrees of underlying coupling between the target sources is depicted in figure 5.3. CCA and mSPoC both assume an existing coupling between modalities and therefor their performance degrades as the coupling becomes weaker.

Finally, figure 5.4 shows the performance of the compared methods as a function of the amount of training data. The performance of mSPoC rises quickly as the amount of training data is increased. For this particular set of simulation parameters, a near-peak performance is reached after about 400 training epochs ($\approx$ 3 minutes of data), after which the performance

still rises but with a much slower rate. CCA and ICA show a similar behavior, but reach their (lower) near-peak performance much quicker.

5.2.2 Real data examples

Fusion of EEG and fNIRS

In this real data example we employ both mSPoC and CCA to simultaneously recorded EEG and fNIRS data, during a study on multimodal BCI, which is described in detail in Fazli et al. (2012). Here we show a re-analysis of Fazli et. al's data.

Experimental Paradigm 14 healthy, right-handed volunteers participated in an experiment which consisted of 2 blocks of motor execution by means of hand gripping. Subjects were asked to close and open their hands with a frequency of 1 Hz for 4 seconds per trial followed by a break of 10.5 ± 1.5 s. An arrow on a display indicated whether to use their left or right hand. They performed 24 trials per condition and block in randomized order resulting in 48 trials per condition.

Data Acquisition During the task EEG and NIRS was simultaneous recorded. With the NIRS-System (NIRScout 8-16, NIRx Medizintechnik GmbH, Germany) 24 channels located mostly around the motor areas as well as frontal and parietal areas were measured at $f_{\text{NIRS}} = 6.25$ Hz sampling frequency. EEG measurement (BrainAmp by Brain Products, Munich, Germany) was performed using 37 Ag/AgCl electrodes with a sampling rate of $f_{\text{EEG}} = 5$ kHz and downsampled to 100 Hz. Sensors of both systems were integrated into a standard EEG cap, resulting in distances between light source and light detectors between 2-3 centimeters.

Data Analysis Raw NIRS data was transformed to concentration changes of HbO and HbR with the modified Beer-Lambert law Delpy et al. (1988); Kocsis et al. (2006) and band-pass filtered leaving frequencies between 0.025 and 0.25 Hz where the hemodynamic response is expected. EEG data was restricted to a subject-specific frequency band that was used by (Fazli et al., 2012). For the vast majority of subjects it was a narrow sub-band of the alpha or beta band range, which is known to reveal responses for motor execution (Ritter et al., 2009). Artifactual NIRS channels were excluded manually via visual inspection of the raw signals and the power spectrum. The band-pass filtered NIRS signals were sub-sampled such that their new sampling rate was 1 Hz.

The EEG data was divided up into non-overlapping epochs of 1 second length, which matched the temporal resolution of the re-sampled fNIRS signals. Here we employed a version of CCA that is also able to account for a temporal filtering of the projection of the EEG power, namely temporal CCA (Bießmann et al., 2009). For the mSPoC analysis we computed EEG covariance matrices within each epoch, while for the CCA analysis a channel- and epoch-wise Fourier transformation was conducted in order to extract the band

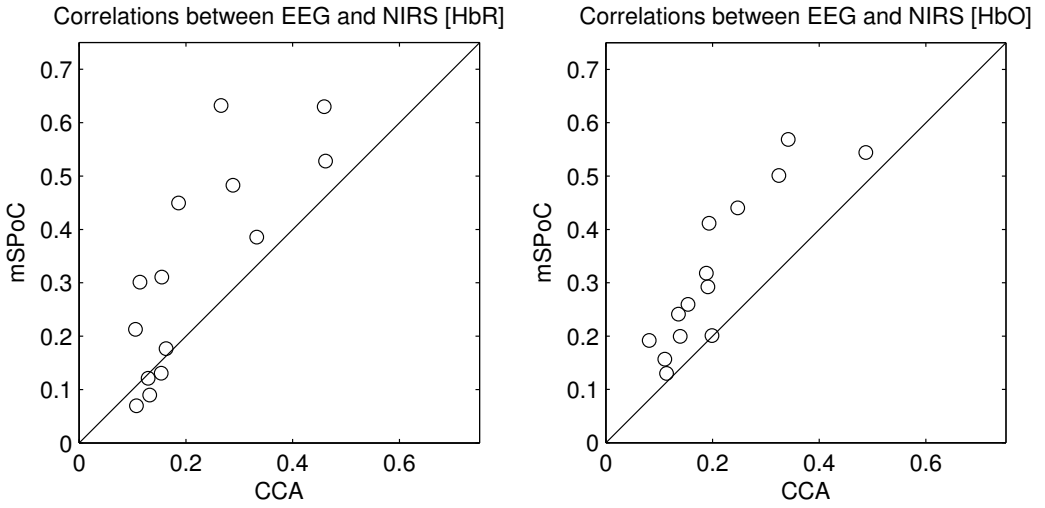


Figure 5.5: Cross-validated correlations between EEG and NIRS (left HbR, right HbO) in the motor execution task for each subject. Results of mSPoC and CCA are compared. Each point corresponds to the correlations obtained for the first set of $\mathbf{w_x}$, $\mathbf{w_y}$, and $\mathbf{w}_\tau$ from a single subject by CCA (x-axis) and mSPoC (y-axis).

power of interest. This yielded a time series of EEG covariance matrices for mSPoC and a multivariate power time series for CCA.

In order to avoid over-fitting and to test the generalization ability of the obtained mSPoC and CCA components, we employed a chronological 5-fold cross-validation.

Results Figure 5.5 shows the results of applying mSPoC and CCA to the simultaneously measured EEG and NIRS data of 14 subjects. We plot the cross-validated correlation values of the first set of components for each subject in a scatter plot. This way one can easily compare the performance of mSPoC (y-axis) and CCA (x-axis) for each subject (single point in the scatter plot). Each point above the diagonal indicates that higher correlations were found with mSPoC as compared to CCA. The scatter plots indicate that mSPoC achieves higher correlations in both analysis settings, i.e. when the convolved band power of EEG components is correlated with NIRS components (HbR and HbO).

We find that mSPoC is able to extract motor-related sources from the two imaging modalities. This is illustrated in figure 5.6, which shows the EEG source patterns of one component set projected on the scalp and the temporal filter for one exemplary subject (VPean) for each motor class separately. Plots in the top row are obtained from left hand movement trials and plots shown in the bottom row are obtained from right hand movement trials. The spatial pattern of the extracted EEG source in the top row of the figure is consistent with a dipole in motor-related areas of the right hemisphere. The NIRS pattern also has its largest weights located over motor areas on the right side of the scalp, which nicely reflects

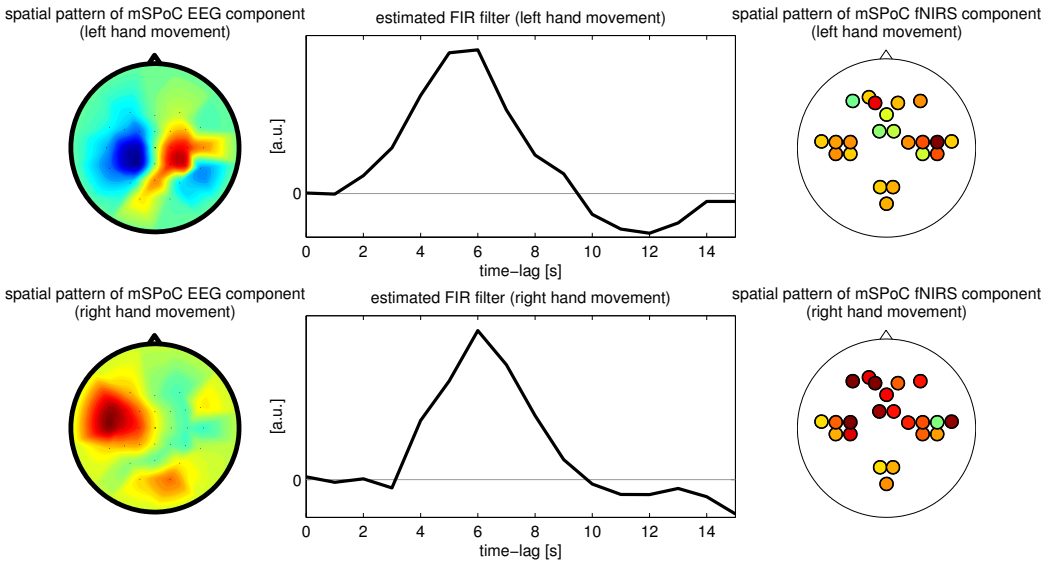


Figure 5.6: Exemplary results for one subject (VPean) as derived by mSPoC. The scalp-plots on the left side show the EEG pattern that corresponds to the obtained filter $\mathbf{w_x}$. In the middle plot we show the temporal filter for the EEG power of the component shown left. The rightmost scalp-plots depict the spatial pattern that corresponds to the filter $\mathbf{w_y}$, i.e. the NIRS patterns. The top row shows the results for applying mSPoC to left hand movement trials, while in the bottom row results for right hand movement trials are shown.

the neuro-physiology for left hand movements. The same is true for patterns obtained on the right hand movement trials. The EEG pattern is located over motor areas of the left hemisphere, which is consistent with the center of mass of the NIRS pattern. The shape of the derived temporal filter of both motor classes is in accordance with the HRF reported in the literature (Boynton et al., 1996). The filter functions realize a temporal low-pass filter with a time shift of 5 to 8 seconds.

Fusion of EEG and fMRI

In this second real world example we illustrate the application of mSPoC for the fusion of a multimodal dataset of simultaneously recorded EEG and fMRI.

Experimental Paradigms One subject was placed in an MR scanner and instructed to squeeze a soft ball five consecutive times with a frequency of 1 to 2 Hz each time an auditory brief tone (frequency = 600 Hz; duration = 200 ms) was presented. The interval between each trial of hand movements was varied from 20 s to 25 s (mean = 22.5 s) and the subjects were instructed to keep their eyes open. This part of the experiment comprised three runs, each consisting of 16 cued movements and each lasting around six minutes. In the following we refer to the data of this part of the experiment as the *motor data*.

Additionally, one block of "eyes open"/"eyes closed" measurements was conducted. In this block, the subject was asked to relax and switch between eyes open and eyes closed every time an auditory cue was played. The total run time of this block was 7 minutes with switching cues occurring every 30 seconds. The data from this part of the experiment is referred to as the *eyes open/closed data*.

Data Aqcuisition Functional MRI data were recorded with a 3 T MRI System while the subject performed the motor task. Whole-brain EPI images were recorded every 2 s (30 axial slices; TE = 30 ms; flip angle = 90°; field of view = 192 x 192 mm; voxel size = 3 x 3 x 4 mm, interslice gap = 1 mm). To minimize artifacts in the EEG, the internal MRI ventilation system was disabled and the helium pump was switched off (Ritter and Villringer, 2006; Nierhaus et al., 2013). EEG was simultaneously recorded inside the MR-scanner (31 scalp channels, arranged according to the International 10—20 System, plus one electrocardiogram channel (ECG) at participants' back).

Data analysis MRI data were preprocessed using SPM 8 (Wellcome Trust Centre for Neuroimaging, UCL, London, UK). Functional volumes were slice time-corrected using the middle slice as reference, realigned to the first image of the time series, and corrected for movement-induced image distortions (6-parameter rigid body affine realignment). Then, images were normalized to the Montreal Neurological Institute (MNI) EPI-template. Note that no spatial smoothing was applied. From the $53 \cdot 63 \cdot 46 \approx 150000$ voxels that constitute a single scan volume, a subset of approximately 75000 voxels was selected based on a gray matter template provided by the SPM software. Voxels for which the gray matter probability was larger than 0.1 were kept while the remaining voxels were discarded. In

order to remove drifts the resulting fMRI time-series were high-pass filtered with a cutoff frequency of 0.0083 Hz (corresponding to a period of 120 s). Thereafter, the fMRI was upsampled to 1 Hz sampling rate. Finally, linear kernel PCA was applied to reduce the dimensionality of the fMRI to a subset of components that together explained 95% of the total variance.

The EEG data was preprocessed using a Matlab toolbox presented in Liu et al. (2012). Processing steps performed with this toolbox included removal of MR gradient artifacts as well as removal of the ballistocardiographic artifact. Further EEG preprocessing was performed using custom Matlab code and included band-pass filtering and dimensionality reduction. Hand movements are expected to modulate the previously mentioned sensorimotor rhythm (SMR). EEG-only analysis revealed the strongest modulation in the beta band for this subject during the motor execution blocks. Thus for mSPoC analysis, the EEG motor data were bandpass filtered between 16 Hz and 22 Hz (*low beta*) as well as between 21 Hz and 30 Hz (*high beta*). The eyes open/closed data were bandpass filtered between 8 Hz and 12 Hz, which corresponds to the well known alpha rhythm. Furthermore, a dimensionality reduction based on spatio-spectral decomposition (SSD) (Nikulin et al., 2011; Haufe et al., 2014a) was performed to increase the SNR on the selected band by retaining 20 SSD components. Consecutive and non-overlapping EEG epochs were defined based on the scan-triggers send by the fMRI scanner. This resulted in epochs of 1 s length temporally aligned with the up-sampled fMRI time-series.

After mSPoC analysis, spatial activation patterns were computed for EEG and fMRI according to Eq. (2.4). The fMRI activation map was z-transformed and values $|z| \leq 3.5$ were set to zero. All reported coordinates correspond to the anatomical MNI space and the probability maps of the Anatomy Toolbox (version 1.7; (Eickhoff et al., 2007)) were used for assignment of clusters in the fMRI activation pattern to their underlying Brodmann areas (BAs). For illustration, the fMRI pattern was superimposed on an anatomical template as provided in MRIcroGL (http://www.mccauslandcenter.sc.edu/mricrogl/).

Finally, in order to assess the physiological meaningfulness of the obtained components, the brain region generating the first EEG mSPoC component from the motor data was source-localized using the multiple signal classification approach (MUSIC Schmidt, 1986). The MUSIC algorithm subsequently scanned through 74,661 dipole locations located on the tesselated cortical surface, and measured, using the corresponding part of the lead field, to what extent a single dipole at each location can explain the activation pattern $\mathbf{a_x}$.

Results Figures 5.7 and 5.8 show the first mSPoC component pairs obtained from the motor data where the EEG has been bandpass filtered for the beta range (low and high beta, respectively). Depicted are the spatial activation patterns of the EEG component and the activation patterns of their corresponding fMRI components, $\mathbf{a_x}$ and $\mathbf{a_y}$, respectively. Furthermore, the figure shows the cross-correlation between the log-bandpower of the EEG component and the time-course of its corresponding fMRI component. In addition the estimated FIR filter coefficients of the vector $\mathbf{w_x}$ are plotted. Finally, the event-related (de-)synchronization of the EEG component is shown using a spectrogram plot and the result of the MUSIC scan (i.e. distributed dipole fit) as shown in the figures as well.

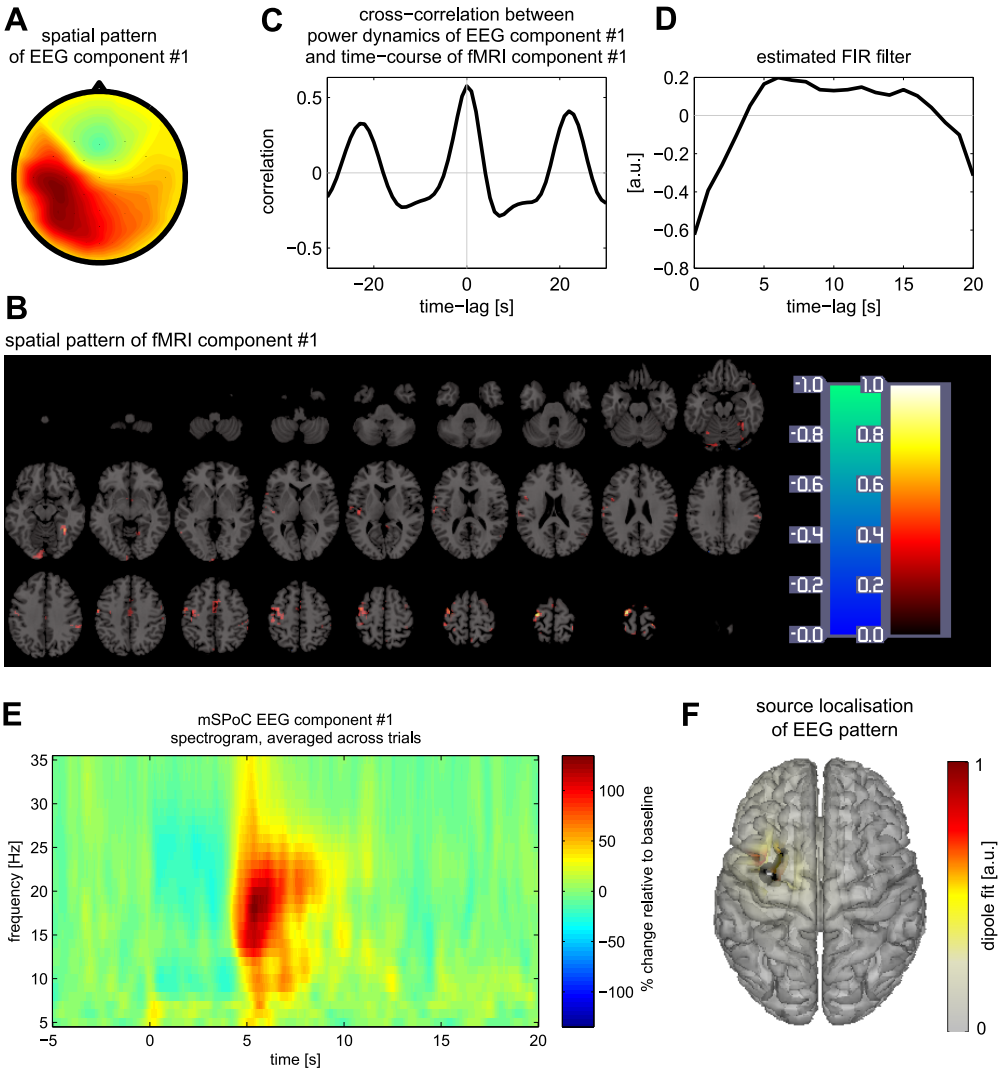


Figure 5.7: Results of mSPoC analysis of motor data: *low-beta*. **A:** Spatial activation pattern $\mathbf{a_x}$ of mSPoC EEG component. **B:** Spatial activation pattern of the fMRI component plotted on a template brain. Shown are 27 equidistant transverse slices through the 3D activation pattern. **C:** Cross-correlation between time-course of *low-beta* power of this EEG component and the extracted BOLD time-course of the corresponding fMRI component. **D:** Coefficients of the estimated FIR filter $\mathbf{w}_\tau$. **E:** Spectrogram (averaged across 45 trials) showing event-related (de-)synchronization (ERD/ERS) of the extracted EEG signal. **F:** Source localization of the EEG spatial pattern using the MUSIC algorithm (Schmidt, 1986).

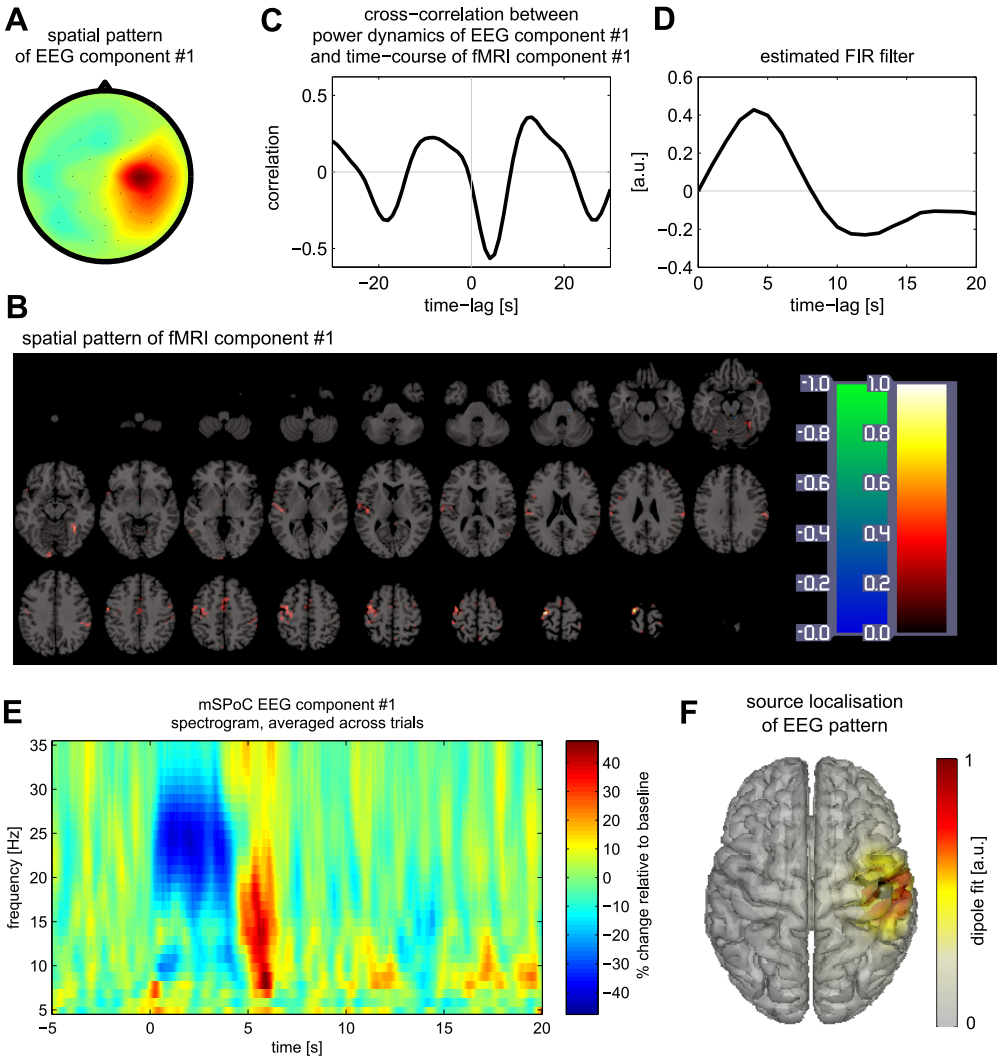


Figure 5.8: Results of mSPoC analysis of motor data: *high-beta*. **A:** Spatial activation pattern $\mathbf{a_x}$ of mSPoC EEG component. **B:** Spatial activation pattern of the fMRI component plotted on a template brain. Shown are 27 equidistant transverse slices through the 3D activation pattern. **C:** Cross-correlation between time-course of *high-beta* power of this EEG component and the extracted BOLD time-course of the corresponding fMRI component. **D:** Coefficients of the estimated FIR filter $\mathbf{w}_\tau$. **E:** Spectrogram (averaged across 45 trials) showing event-related (de-)synchronization (ERD/ERS) of the extracted EEG signal. **F:** Source localization of the EEG spatial pattern using the MUSIC algorithm (Schmidt, 1986).

We shall discuss the extracted low beta component in more detail now. The largest activation cluster in the fMRI activation pattern (k = 243 voxels) covered the left sensorimotor cortex, with maximal activations in the premotor cortex (BA 6; x = -24, y= -13, z= 73), primary motor cortex (BA 4p; x= -33, y = -22, z = 55), and primary somatosensory cortex (BA1; x = -45, y = -34, z = 64). This is in line with previous studies showing that Rolandic rhythm strength is (inversely) correlated with activity in motor and somatosensory cortex (Ritter et al., 2009). Note however that in this particular case the highest correlation between EEG beta power and BOLD signal is actually positive. This is explained by the spectrogram plot in the lower left panel of the figure. The plot shows that in addition to a slight decrease in beta power relative to baseline, there is a large increase (more than 100% relative to baseline) after about four to five seconds. This effect is called *beta rebound* and it is found to occur after executed as well as imagined movements (Pfurtscheller et al., 1996, 2005; Jurkiewicz et al., 2006). In this subject, the peak of the beta rebound coincides with maximal BOLD signal activation, which is evident from the maximum in the cross-correlation at lag zero. Hence also the maximum-magnitude weight in the FIR filter at lag zero.

The EEG pattern of the mSPoC component was best approximated by a dipole in the left primary motor cortex (BA 4p; MNI coordinates x = -27, y= -4, z= 51) pointing tangentially towards the front right (normalized orientation x = 0.78, y = 0.62, z = 0.11). The EEG source was thus estimated to be less than 2 cm away from the corresponding fMRI activation in primary motor cortex. The best-fitting dipole almost perfectly explained the mSPoC activation pattern ($v = 98\,\%$ explained variance). The panel in the lower right of the figure depicts the variance v (scaled as $v/(1-v)$) explained by a dipole in each location on the cortical surface, where the location and orientation of the best-fitting dipole is indicated by a black arrow.

Interestingly, the EEG pattern of the first mSPoC component obtained for the high beta band-pass filter settings (see Figure 5.8) is consistent with a dipolar source in the right hemisphere of the brain, and thus ipsilateral to the hand that was used for movement. Cross-correlating the beta band time-course of the EEG component with the temporal activation of the corresponding fMRI component reveals near zero correlation at zero lag and maximal (inverse) correlation at a time lag of 4 seconds. This behavior is reflected in the estimated FIR filter coefficients, which have their maximum at a time lag of 4 seconds as well.

The spectrogram of the EEG component's time course shows a strong decrease in high beta power immediately after movement onset and only very little rebound in high beta. A post-movement beta rebound can be observed in the lower beta range of this component as well, yet the magnitude of this effect is only half of what was observed for the contral-lateral EEG component.

Figure 5.9 shows the results obtained from the eyes open/closed dataset. Similar to the figures from the motordata, this figure depicts EEG and fMRI activation patterns, the cross-correlation between EEG component bandpower (here from the alpha band) and the coefficients of the estimated FIR filter. In addition, this figure also shows the time-courses of the EEG component bandpower (with and without application of FIR filter) as well as the time-course of the extracted fMRI component.

The spatial activation patterns of the EEG and fMRI component indicate occipital regions as the generators of the extracted signals. As for the temporal relation between the modalities, the strongest correlation between band power of the EEG component and the extracted fMRI signal is observed at a time lag of four seconds. Similar to what we have seen before, the FIR filter coefficients reflect the delay in maximal coupling. The bottom panel of Figure 5.9 shows the temporal co-modulation of extracted EEG and fMRI. Both signals are clearly modulated by the alternation between blocks of eyes open and eyes closed, alternating every 30 seconds.

In summary, we find that mSPoC applied to simultaneously recorded EEG and fMRI extracted meaningful components the spatial, spectral and temporal dynamics of which were in line with previous findings and/or were expected given what is known in the literature. Thereby, these real data experiments further validated the use of mSPoC for the fusion of multimodal neuroimaging data.

5.3 Discussion

In this chapter we presented a novel technique for multimodal integration with emphasis on bandpower dynamics. The objective function we proposed is designed to reflect the generative models of EEG and hemodynamic signals. The new analysis method is called mSPoC and we compared it to a state-of-the-art competitor CCA. The results show that the performance of mSPoC exceeds that of CCA on simulated as well as real world data.

The proposed optimization scheme is efficient and fast enough to allow an application to real experimental data. However, we expect that there might still be room for improvement. There are a number of parameters that may be subjected to further optimization in order to reach better performance, for example the length of the epoch-window or the frequency band used to band-pass filter the EEG. Also the actual optimization algorithm may still be improved.

Related methods In Dähne et al. (2013) we have shown that the mSPoC objective can be expressed in terms of tensor factorization. There are other tensor factorization approaches (also called multi-way analysis) to either EEG power alone (Lee et al., 2007; Mørup et al., 2008) or to EEG power in conjunction to hemodynamics (Martínez-Montes et al., 2004). In these approaches, however, EEG band power is computed in sensor space as part of the preprocessing and then added as a single dimension into a tensor or multi-way array. In our approach, the nonlinearity (i.e. estimation EEG band power, here approximated by squaring) is directly incorporated into the objective function and into the required data structure as the EEG data enters the objective function as a time-series of covariance matrices. This is a necessary step in order to represent the EEG *source* power rather than the EEG *channel* power.

Physiological plausibility Our novel framework corresponds to a widely accepted generative model for the electro-physics of the EEG and the hemodynamics of neural sources (Nunez and Srinivasan, 2005). Therefore, our results are interpretable within this model. This is a

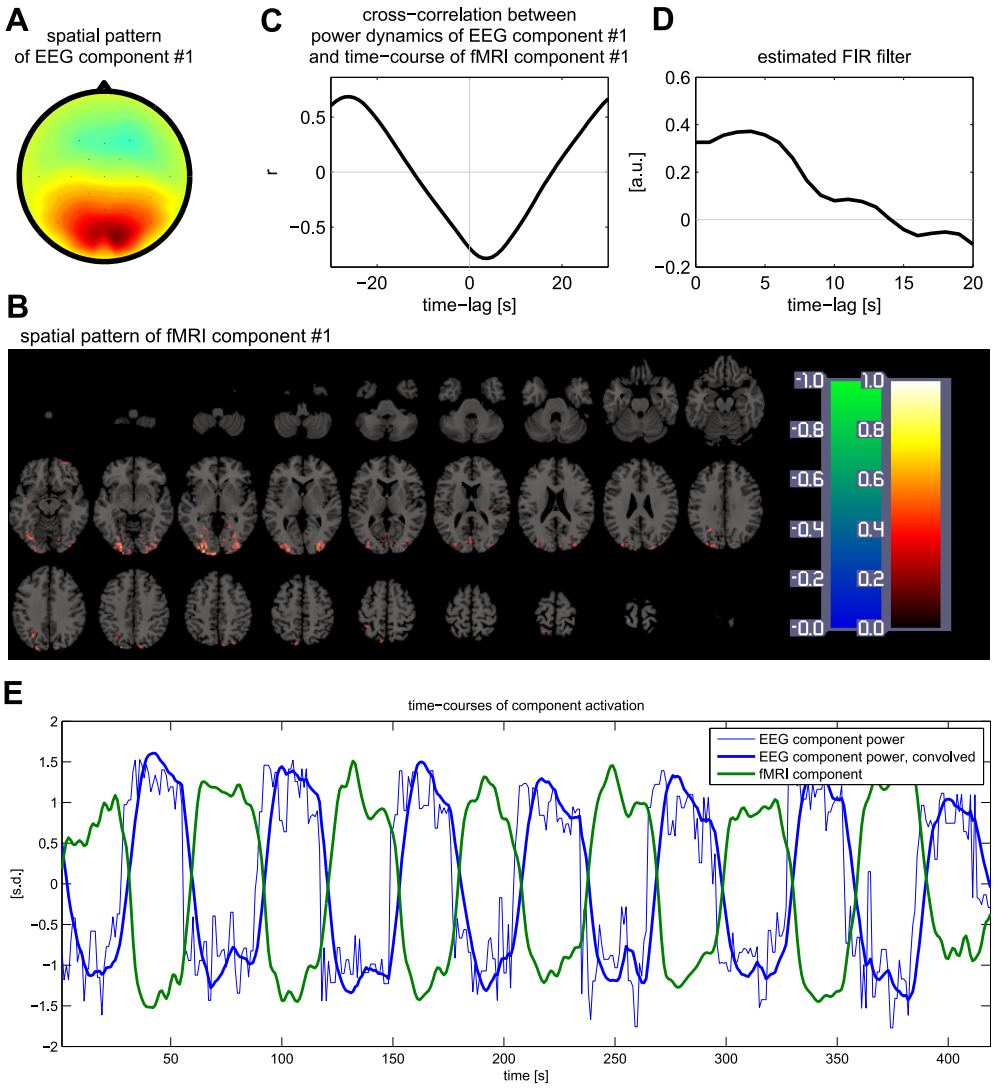


Figure 5.9: Results of mSPoC analysis of eyes open/closed data: *alpha*. **A:** Spatial activation pattern $\mathbf{a_x}$ of mSPoC EEG component. **B:** Spatial activation pattern of the fMRI component plotted on a template brain. Shown are 27 equidistant transverse slices through the 3D activation pattern. **C:** Cross-correlation between time-course of *alpha* power of this EEG component and the extracted BOLD time-course of the corresponding fMRI component. **D:** Coefficients of the estimated FIR filter $\mathbf{w}_\tau$. **E:** Time-courses of the mSPoC EEG component alpha power (with and without subsequent FIR filtering) and the mSPoC fMRI component activation. Switching between eyes open and eyes closed was cued every 30 seconds, starting with eyes open.

particular difference to CCA- or PLS-based approaches for example. The spatial filters for the EEG band power obtained by our method are invertible to yield spatial patterns which correspond to the underlying sources. These EEG patterns can in principle be subjected to source localization algorithms (e.g. Baillet et al., 2001; Pascual-Marqui et al., 1994; Schmidt, 1986; Gramfort et al., 2013; Haufe et al., 2011), which is not true for patterns obtained from CCA or PLS, because these methods act on band power representations of the data.

We find physiologically plausible temporal coupling dynamics, which are in good accordance with previous studies (Boynton et al., 1996; Murayama et al., 2010; Bießmann et al., 2009). Also the spatial patterns extracted by mSPoC are in line with established knowledge about the spatial location of the dipoles relevant for the experimental paradigm investigated here. A more thorough empirical validation with respect to the hemodynamic coupling between EEG and NIRS has been made possible with our method and is highly desirable. Note that our presented modeling approach does *not* assume the dynamics of the hemodynamic activation to be identical for all positions in the brain, which would imply the existence of location-independent *canonical hemodynamic response function* (HRF). Instead, by optimizing $\mathbf{w}_\tau$ anew together with each component pair $\mathbf{w_x}$ and $\mathbf{w_y}$, our mSPoC approach explicitly models a potentially space-varying and non-instantaneous coupling between EEG bandpower dynamics and fMRI activations. This is in line with the known variability of the HRF across space and subjects (Handwerker et al., 2004; Pedregosa et al., 2015; Bießmann et al., 2012).

Conclusion In this chapter we have addressed the following question: how can we correlate two raw multimodal multivariate signals that are spatially and temporally sampled differently and that are coupled through a nonlinear transformation? Our answer to this abstract question allows to find *unsupervised* projections, that discover unknown correlations between spectral power and a hemodynamic response. Furthermore, mSPoC represents an important contribution towards data-driven multimodal data integration with a minimal set of assumptions. Thus, it has the potential to enable insights about complex cognitive processes in future studies.

Chapter 6

Canonical Source Power Co-modulation

In the previous two chapters we have addressed the optimal extraction of neuronal oscillations which exhibit band-power that are correlated with i) univariate target signals such as stimulus- or cognitive variables in the case of SPoC, and ii) multivariate target signals such as signals from a different measurement modality in the case of mSPoC. In this chapter, we continue this line of research, and investigate the extraction of pairs of neuronal sources with band-power dynamics that are correlated with *each other*. This type of functional connectivity is called amplitude-amplitude coupling (AAC) or power-to-power coupling. *Intra*-subject power-to-power coupling may be observed as a (positive or negative) correlation of the power dynamics of two neural processes within the same frequency band or across frequency bands, and is hypothesized to mediate long as well as short range information transfer between brain areas (FitzGerald et al., 2013; Furl et al., 2014). In studies such as social neuroscience experiments investigating *inter*-subject interaction, power-to-power coupling between subjects can reveal leader/follower relations (Sänger et al., 2013). Additionally, the analysis of inter-subject power-to-power coupling could identify brain responses to potentially complex external stimuli purely based on the assumption that the stimulus-induced power dynamics are consistent across subjects. This is the idea of *hyperscanning* (Hasson et al., 2004; Montague et al., 2002).

In this chapter we present a method for the extraction of components from EEG/MEG recordings that exhibit power-to-power coupling. We refer to this method as *canonical Source Power Correlation* analysis (cSPoC). We demonstrate the utility of cSPoC in simulations as well as in a number of relevant application scenarios, each involving real EEG data of multiple subjects. These real-world scenarios are

- the isolation of oscillatory brain responses modulated by a complex auditory stimulus, based on the idea that the bandpower dynamics of these responses are similar across multiple repetitions of the stimulus and for multiple subjects.

- the blind recovery of pairs of alpha-band oscillations and their corresponding first higher harmonics in the beta-band, based on the idea that the various harmonics of the alpha rhythm are different aspects of the same neural process and should thus exhibit strong envelope correlations.

- the extraction of oscillatory source components from simultaneously recorded EEG and MEG that reflect ongoing dynamics of the brain during resting state.

Note that for all application scenarios considered the ground truth is either known or at least a strong hypothesis about the expected outcome exists. Therefore, all our analyses serve as real-world verifications for the effectiveness of cSPoC.

Our algorithm is called *canonical* source power correlation analysis, because it bears some conceptual similarity with canonical correlation analysis (CCA, Hotelling, 1936). CCA and its variants are powerful tools for the extraction of correlated components from two multivariate data sets. Both CCA and cSPoC seek linear projections of two data sets in order to maximize a correlation coefficient defined on the projected data (here the component time-courses). However, in case of CCA, the correlation is defined between the component time courses directly, whereas in cSPoC the correlation is defined between nonlinear features of the components, namely their envelopes. While it is conceivable to use a two-step approach, i. e., first compute the power dynamics (e.g. envelopes) of the raw channel data of the two data sets and then apply CCA, such an approach does not lead to an accurate inversion of the generative model Eq. (2.1) (see also Dähne et al., 2013, and previous chapters of this thesis). The reason for this is that the two operations (i) source extraction through linear projection and (ii) nonlinear processing (computation of the envelope) do not commute. By design, cSPoC computes power dynamics on the component level, i.e. *after* projecting the data onto the extraction filters, and thus performs both operation in the order implied by the generative model. Figure 6.1 graphically illustrates the fundamental difference between CCA and cSPoC.

6.1 The cSPoC algorithm

Before we begin the derivation of cSPoC let us briefly review the necessary notation. Here we want to find pairs of oscillatory neural sources from electrophysiological measurements based on the assumption that the power dynamics of these sources are correlated. Thus, we require the presence of a second set of measured oscillatory data. In order not to confuse the notation with previous sections, we denote the second oscillatory data set by the multivariate variable $\acute{\mathbf{x}}(t) \in \mathbb{R}^{N_{\acute{x}}}$ with sources $\mathbf{s}_{\acute{\mathbf{x}}}(t) \in \mathbb{R}^{K_{\acute{x}}}$, mixing coefficients $\mathbf{A}_{\acute{\mathbf{x}}} \in \mathbb{R}^{N_{\acute{x}} \times K_{\acute{x}}}$ and noise $\epsilon_{\acute{\mathbf{x}}}(t) \in \mathbb{R}^{K_{\acute{x}}}$. The datasets $\mathbf{x}$ and $\acute{\mathbf{x}}$ are assumed to be related by $K < \min(K_x, K_{\acute{x}})$ pairs of source processes among the rows of $\mathbf{s}_{\mathbf{x}}$ and $\mathbf{s}_{\acute{\mathbf{x}}}$, whose power dynamics are linearly correlated. As a prerequisite for this, both datasets must be synchronized; that is, share a common time or sample index $t \in \{1, \cdots, T\}$. This synchronization could be achieved by performing concurrent (e. g., EEG) recordings of two different subjects, by performing concurrent recordings of a single subject in two different measurement modalities (e. g., EEG and ECoG), by performing two recordings of a single subject (or single recordings of two different subjects) under identical stimulation, by creating two independent spectral representations of a single measurement through spectral filtering in two different frequency bands, or simply by using the same dataset for both $\mathbf{x}$ and $\acute{\mathbf{x}}$, and so on.

6.1.1 Extracting a single source component pair

First we describe the extraction of a single envelope-coupled source component pair. Subsequent pairs can then be obtained using the deflation scheme described in section A.3.

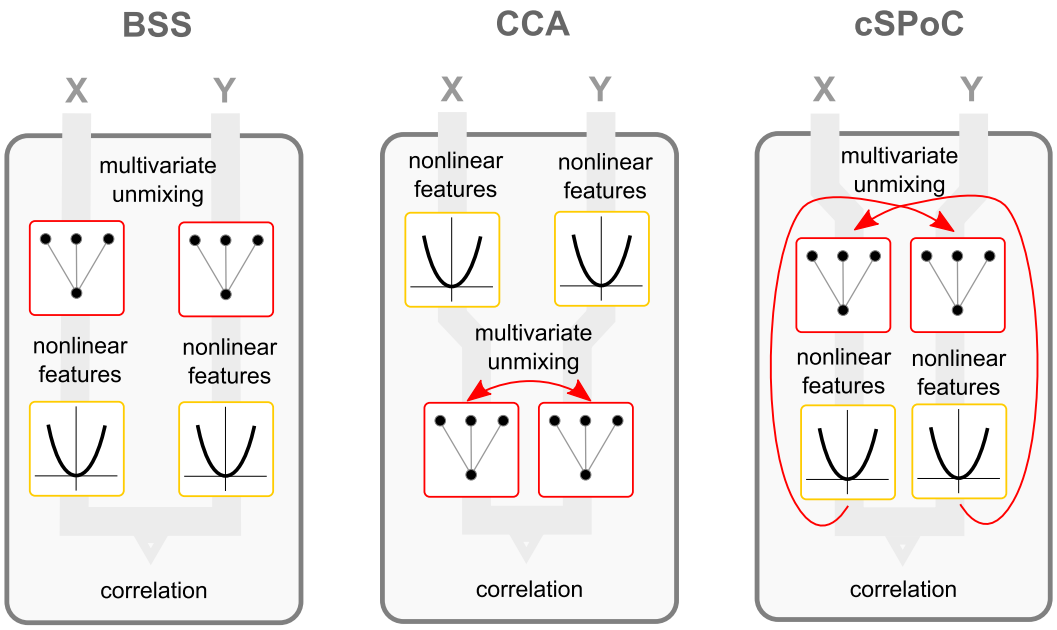


Figure 6.1: Illustration of different approaches extract components for power-to-power coupling. The input to all three approaches two datasets of multivariate and bandpass filtered EEG/MEG data $\mathbf{x}$ and $\acute{\mathbf{x}}$. Processing steps are organized from top to bottom. **Left:** An approach that is based on blind source separation (BSS) methods that are applied to each modality independently. This approach is in line with the generative model we assume for both modalities. However, the modality-specific BSS methods do not inform each other during the decomposition and thus potentially miss out on available information. Furthermore, after the decomposition step, a pair-wise matching of components with similar bandpower dynamics is required. **Middle:** An approach based on canonical correlation analysis (CCA). In this approach, the decomposition into components is informed by the respective other dataset and an additional pair-wise matching of components is not required. The crucial drawback of this approach is that bandpower has to be computed on channel-level, which leads to a number of shortcomings that have been outlined in chapter 3. **Right:** Our proposed approach called canonical source power co-modulation (cSPoC) that combines advantages of the two other approaches (conformity with the generative model as well as maximal use of cross-dataset information) while avoiding their shortcomings.

Let $\mathbf{w}_\mathbf{x} \in \mathbb{R}^{N_x}$ and $\mathbf{w}_{\acute{\mathbf{x}}} \in \mathbb{R}^{N_{\acute{x}}}$ be extraction filters for a single source pair $\mathbf{s}_\mathbf{x} = \mathbf{w}_\mathbf{x}^\top \mathbf{x}$ and $s_{\acute{\mathbf{x}}} = \mathbf{w}_{\acute{\mathbf{x}}}^\top \acute{\mathbf{x}}$. The instantaneous amplitudes (envelopes) of these sources are given by

$$\phi\left[s_\mathbf{x}\right] = \sqrt{(\mathbf{w}_\mathbf{x}^\top \mathbf{x})^2 + (\mathbf{w}_\mathbf{x}^\top \mathrm{H}[\mathbf{x}])^2} \tag{6.1}$$

$$\phi\left[s_{\acute{\mathbf{x}}}\right] = \sqrt{(\mathbf{w}_{\acute{\mathbf{x}}}^\top \acute{\mathbf{x}})^2 + (\mathbf{w}_{\acute{\mathbf{x}}}^\top \mathrm{H}[\acute{\mathbf{x}}])^2}, \tag{6.2}$$

where $\mathrm{H}[\cdot]$ denotes the Hilbert transform (Barlow, 1993). We use $\phi_\mathbf{x}$ and $\phi_{\acute{\mathbf{x}}}$ as shorthand notations for $\phi\left[\mathbf{w}_\mathbf{x}^\top \mathbf{x}\right]$ and $\phi\left[\mathbf{w}_{\acute{\mathbf{x}}}^\top \acute{\mathbf{x}}\right]$ from here on.

With these definitions, we can express the correlation between the source envelopes as a function of the spatial filters $\mathbf{w}_\mathbf{x}$ and $\mathbf{w}_{\acute{\mathbf{x}}}$. The cSPoC objective function thus reads

$$\max_{\mathbf{w}_\mathbf{x}, \mathbf{w}_{\acute{\mathbf{x}}}} \; c_{\mathrm{sg}} \cdot \mathrm{Corr}[\phi_\mathbf{x}, \phi_{\acute{\mathbf{x}}}] \,, \tag{6.3}$$

where the constant $c_{\mathrm{sg}} \in \{+1, -1\}$ decodes whether positive or negative correlations are desired, and where

$$\mathrm{Corr}[\phi_\mathbf{x}, \phi_{\acute{\mathbf{x}}}] = \frac{\mathrm{Cov}[\phi_\mathbf{x}, \phi_{\acute{\mathbf{x}}}]}{\sqrt{\mathrm{Var}[\phi_\mathbf{x}]\,\mathrm{Var}[\phi_{\acute{\mathbf{x}}}]}} \tag{6.4}$$

$$= \frac{\left\langle \bar{\phi}_\mathbf{x} \bar{\phi}_{\acute{\mathbf{x}}} \right\rangle}{\sqrt{\left\langle (\bar{\phi}_\mathbf{x})^2 \right\rangle \left\langle (\bar{\phi}_{\acute{\mathbf{x}}})^2 \right\rangle}} \tag{6.5}$$

is Pearson's product-moment correlation coefficient with $\langle \cdot \rangle$ denoting average over time and with $\bar{\phi}_\mathbf{x} = \phi_\mathbf{x} - \langle \phi_\mathbf{x} \rangle$ and $\bar{\phi}_{\acute{\mathbf{x}}} = \phi_{\acute{\mathbf{x}}} - \langle \phi_{\acute{\mathbf{x}}} \rangle$.

The objective stated in Eq. (6.3) is a non-convex higher-order nonlinear function of the coefficients $\mathbf{w}_\mathbf{x}$ and $\mathbf{w}_{\acute{\mathbf{x}}}$. We propose to optimize it by means of standard nonlinear optimization techniques such as the limited-memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) algorithm (Nocedal, 1980) implemented in MATLAB's (The Mathworks) `fminunc` routine. After the first pair $\mathbf{w}_\mathbf{x}$ and $\mathbf{w}_{\acute{\mathbf{x}}}$ is found, subsequent pairs can be extracted using the deflation scheme outline in chapter A.3.

The gradient of Eq. (6.3) with respect to $\mathbf{w}_\mathbf{x}$ is given next. For symmetry reasons, the gradient with respect to $\mathbf{w}_{\acute{\mathbf{x}}}$ is obtained analogously and thus not shown here.

$$\frac{\partial\, \mathrm{Corr}[\phi_\mathbf{x}, \phi_{\acute{\mathbf{x}}}]}{\partial \mathbf{w}_\mathbf{x}} = \frac{\partial \frac{\left\langle \bar{\phi}_\mathbf{x} \bar{\phi}_{\acute{\mathbf{x}}} \right\rangle}{\sqrt{\left\langle (\bar{\phi}_\mathbf{x})^2 \right\rangle \left\langle (\bar{\phi}_{\acute{\mathbf{x}}})^2 \right\rangle}}}{\partial \mathbf{w}_\mathbf{x}}$$

$$= \frac{\left\langle \Delta_{\acute{\mathbf{x}}\mathbf{x}} \frac{\partial \bar{\phi}_\mathbf{x}}{\partial \mathbf{w}_\mathbf{x}} \right\rangle}{\sqrt{\left\langle (\bar{\phi}_\mathbf{x})^2 \right\rangle \left\langle (\bar{\phi}_{\acute{\mathbf{x}}})^2 \right\rangle}}, \tag{6.6}$$

with

$$\bar{\phi}_i = \phi_i - \langle \phi_i \rangle \,, \text{ for } i \in \{\mathbf{x}, \acute{\mathbf{x}}\}, \tag{6.7}$$

$$\Delta_{\acute{\mathbf{x}}\mathbf{x}} := \bar{\phi}_{\acute{\mathbf{x}}} - \bar{\phi}_{\mathbf{x}} \frac{\left\langle \bar{\phi}_{\mathbf{x}} \bar{\phi}_{\acute{\mathbf{x}}} \right\rangle}{\left\langle (\bar{\phi}_{\mathbf{x}})^2 \right\rangle}, \tag{6.8}$$

$$\frac{\partial \bar{\phi}_{\mathbf{x}}}{\partial \mathbf{w}_{\mathbf{x}}} = \frac{\partial \phi_{\mathbf{x}}}{\partial \mathbf{w}_{\mathbf{x}}} - \left\langle \frac{\partial \phi_{\mathbf{x}}}{\partial \mathbf{w}_{\mathbf{x}}} \right\rangle. \tag{6.9}$$

For $\phi_{\mathbf{x}} = \sqrt{(\mathbf{w}_{\mathbf{x}}^\top \mathbf{x})^2 + (\mathbf{w}_{\mathbf{x}}^\top \mathrm{H}[\mathbf{x}])^2}$, we have

$$\frac{\partial \phi_{\mathbf{x}}}{\partial \mathbf{w}_{\mathbf{x}}} = \frac{1}{\phi_{\mathbf{x}}} \left((\mathbf{w}_{\mathbf{x}}^\top \mathbf{x})\, \mathbf{x} + (\mathbf{w}_{\mathbf{x}}^\top \mathrm{H}[\mathbf{x}])\, \mathrm{H}[\mathbf{x}] \right), \tag{6.10}$$

with $\mathrm{H}[\mathbf{x}]$ denoting the Hilbert transform of $\mathbf{x}$.

6.1.2 Extensions

Here, we provide a general formalism for integrating some useful extensions into the cSPoC objective. An extension corresponds to a slight modification of the objective function. Thus, for each extension we also provide the necessary adjustments to the gradient. To this end, let us consider to optimize correlations between general functions f of $\phi_{\mathbf{x}}$ and $\phi_{\acute{\mathbf{x}}}$, where f must be differentiable. The modified optimization problem then becomes

$$\max_{\mathbf{w}_{\mathbf{x}}, \mathbf{w}_{\acute{\mathbf{x}}}} \; c_{\mathrm{sg}} \cdot \mathrm{Corr}[f(\phi_{\mathbf{x}}), f(\phi_{\acute{\mathbf{x}}})] \,. \tag{6.11}$$

The partial derivative of this problem with respect to $\mathbf{w}_{\mathbf{x}}$ is given by

$$\begin{aligned} \frac{\partial\, \mathrm{Corr}[f(\phi_{\mathbf{x}}), f(\phi_{\acute{\mathbf{x}}})]}{\partial \mathbf{w}_{\mathbf{x}}} &= \frac{\partial \frac{\left\langle \bar{f}(\phi_{\mathbf{x}}) \bar{f}(\phi_{\acute{\mathbf{x}}}) \right\rangle}{\sqrt{\left\langle \bar{f}(\phi_{\mathbf{x}})^2 \right\rangle \left\langle \bar{f}(\phi_{\acute{\mathbf{x}}})^2 \right\rangle}}}{\partial \mathbf{w}_{\mathbf{x}}} \\ &= \frac{\left\langle \Delta_{\acute{\mathbf{x}}\mathbf{x}} \frac{\partial \bar{f}(\phi_{\mathbf{x}})}{\partial \mathbf{w}_{\mathbf{x}}} \right\rangle}{\sqrt{\left\langle \bar{f}(\phi_{\mathbf{x}})^2 \right\rangle \left\langle \bar{f}(\phi_{\acute{\mathbf{x}}})^2 \right\rangle}} \end{aligned} \tag{6.12}$$

with

$$\bar{f}(\phi_i) := f(\phi_i) - \langle f(\phi_i) \rangle \,, \text{ for } i \in \{\mathbf{x}, \acute{\mathbf{x}}\} \,, \tag{6.13}$$

$$\Delta_{\acute{\mathbf{x}}\mathbf{x}} := \bar{f}(\phi_{\acute{\mathbf{x}}}) - \bar{f}(\phi_{\mathbf{x}}) \frac{\left\langle \bar{f}(\phi_{\mathbf{x}}) \bar{f}(\phi_{\acute{\mathbf{x}}}) \right\rangle}{\left\langle (\bar{f}(\phi_{\mathbf{x}}))^2 \right\rangle}, \tag{6.14}$$

$$\frac{\partial \bar{f}(\phi_{\mathbf{x}})}{\partial \mathbf{w}_{\mathbf{x}}} = \frac{\partial f(\phi)_{\mathbf{x}}}{\partial \mathbf{w}_{\mathbf{x}}} - \left\langle \frac{\partial f(\phi)_{\mathbf{x}}}{\partial \mathbf{w}_{\mathbf{x}}} \right\rangle \,. \tag{6.15}$$

Finally, we have

$$\frac{\partial f(\phi)_{\mathbf{x}}}{\partial \mathbf{w}_{\mathbf{x}}} = \frac{\partial f(\phi_{\mathbf{x}})}{\partial \phi_{\mathbf{x}}} \cdot \frac{\partial \phi_{\mathbf{x}}}{\partial \mathbf{w}_{\mathbf{x}}} \,, \tag{6.16}$$

where $\frac{\partial \phi_{\mathbf{x}}}{\partial \mathbf{w}_{\mathbf{x}}}$ is given in Eq. (6.10) and $\frac{\partial f(\phi_{\mathbf{x}})}{\partial \phi_{\mathbf{x}}}$ depends on the specific choice of the function f and will be given below for some examples.

Log-transformation

The co-modulation of Gaussian distributed variables is completely described by Pearson's correlation coefficient (Lee Rodgers and Nicewander, 1988). However, envelope and band-power signals are not Gaussian distributed, as noted for example in Freyer et al. (2009). Thus, in practice, a logarithmic transformation is often applied to make band-power and envelope signals more Gaussian (e. g., Hipp et al., 2012) and thereby make obtained correlation values more meaningful.

In order to explicitly optimize the correlation between log-envelopes, the cSPoC objective function can be modified by choosing the function f to be the natural logarithm. Then we have

$$f_{\log}(\phi_{\mathbf{x}}) := \log(\phi_{\mathbf{x}}) \,, \tag{6.17}$$

and

$$\frac{\partial f_{\log}(\phi_{\mathbf{x}})}{\partial \phi_{\mathbf{x}}} = \frac{1}{\phi_{\mathbf{x}}} \,. \tag{6.18}$$

Averaging within Epochs

In some scenarios it might be beneficial to average the envelope of the sources across short time intervals. We refer to these intervals as epochs and denote them by the index e. Possible reasons for averaging envelopes within epochs may include (i) suppression of noise, (ii) reduction of the sampling rate of the envelope, or (iii) the data may have an epoch (or trial) structure already and the envelope dynamics within trials are not the subject of study but common dynamics across trials are. In this case, the cost function has to be modified, such that $\phi_{\mathbf{x}}$ and $\phi_{\dot{\mathbf{x}}}$ are to be correlated across the epoch index e rather than across the original time index t.

Let T_e be the set of all time indices t that are in the epoch e. We define

$$f_e(\phi_{\mathbf{x}}) := \frac{1}{|T_e|} \sum_{t \in T_e} \phi(t) \,, \tag{6.19}$$

where $|T_e|$ denotes the cardinality of the set T_e. Note that $f_e(\phi_{\mathbf{x}})$ now has a sampling rate of epochs, rather than the original time index t. The corresponding partial derivative is given by

$$\frac{\partial f_e(\phi_{\mathbf{x}})}{\partial \mathbf{w}_{\mathbf{x}}} = \frac{1}{|T_e|} \sum_{t \in T_e} \frac{\partial \phi(t)}{\mathbf{w}_{\mathbf{x}}} \,. \tag{6.20}$$

Time-delayed Coupling

The assumed coupling between envelopes may not necessarily be instantaneous. Therefore, several authors have modeled non-instantaneous coupling dynamics explicitly using temporal delay operators (Bießmann et al., 2009; Dähne et al., 2013; Campi et al., 2013). We adopt this idea by applying a temporal filter $\mathbf{w}_\tau \in \mathbb{R}^{N_\tau}$ to one or both of the envelopes, where N_τ is the number of time-lags considered. The temporal filter $\mathbf{w}_\tau$ is estimated jointly with the

filters $\mathbf{w_x}$ and $\mathbf{w_{\acute{x}}}$, and models temporal delays as well as convolutive interactions between envelopes.

Let $\{\tau_i\}$, with $i \in \{1, \cdots, N_\tau\}$, denote a list of suitable time (or epoch) lags. Then we define

$$f_\tau(\phi_\mathbf{x})(t) := \sum_i^{N_\tau} (\mathbf{w}_\tau)_i \, \phi_\mathbf{x}(t - \tau_i), \tag{6.21}$$

where $(\mathbf{w}_\tau)_i$ denotes the i-th element of the vector $\mathbf{w}_\tau$. In this formulation, $\mathbf{w}_\tau$ becomes a parameter vector over which one can optimize in the same way as over $\mathbf{w_x}$ and $\mathbf{w_y}$. The partial derivatives for this extension are

$$\frac{\partial f_\tau(\phi_\mathbf{x})}{\partial \mathbf{w_x}} = \sum_i^{N_\tau} (\mathbf{w}_\tau)_i \, \frac{\partial \phi_\mathbf{x}(t - \tau_i)}{\partial \mathbf{w_x}} \tag{6.22}$$

$$\frac{\partial f_\tau(\phi_\mathbf{x})}{\partial (\mathbf{w}_\tau)_i} = \phi_\mathbf{x}(t - \tau_i) \; . \tag{6.23}$$

Note that envelopes vary on a slower time scale than the carrier signals. Thus it is useful to combine the last two proposed extensions such that during the optimization $\phi_\mathbf{x}$ is downsampled by averaging within short time intervals before applying the temporal filter $\mathbf{w}_\tau$. This greatly reduces the number of coefficients in $\mathbf{w}_\tau$.

Multiple Datasets

Finally, cSPoC is easily extended to settings in which envelope coupling between N > 2 data sets is suspected. Let $\mathbf{x}^n$ denote the $n-$th data set for $n \in \{1, \cdots, N\}$. In this case we seek a set of N weight vectors $\mathbf{w}^n$. One straightforward way of generalizing correlation coefficient to N > 2 data sets is to optimize the sum of pair-wise correlations. This approach, which has been proposed by Kettenring (1971) in the context of CCA, can be easily adopted also in cSPoC. See Appendix B for more details.

6.1.3 Practical details

In order to ensure meaningful results, care has to be taken in cases where the two data sets $\mathbf{x}$ and $\acute{\mathbf{x}}$ stem from same measurement and cSPoC is set to maximize envelope correlations between the projections $\mathbf{w}_\mathbf{x}^\top \mathbf{x}$ and $\mathbf{w}_{\acute{\mathbf{x}}}^\top \acute{\mathbf{x}}$. If $\mathbf{x}$ and $\acute{\mathbf{x}}$ are identical, the trivial solution to maximizing the objective stated in Eq. (6.3) is given by $\mathbf{w_x} = \mathbf{w_{\acute{x}}} = \mathbf{w}$, which means that any pair of spatial filters will be optimal as long as they are identical. The trivial solution can be avoided by imposing additional constraints on $\mathbf{w_x}$ and $\mathbf{w_{\acute{x}}}$ or, alternatively, by ensuring that $\mathbf{x}$ and $\acute{\mathbf{x}}$ are sufficiently different. The latter can be achieved by band-pass filtering in different frequency bands, for example. Note that the problem of trivial solutions does not arise if $\mathbf{x}$ and $\acute{\mathbf{x}}$ are identical and the objective is to be minimized, i.e. if negative envelope correlations are investigated.

The objective function optimized by cSPoC is non-convex, which means that there is the potential for the optimization algorithm to end up in local optima. In practice, we overcome

this issue by restarting the optimization with different, randomly selected initiations. We then use the solution that yielded the highest envelope correlation on the part of the data that was used to fit the model and apply it to the validation data. For current computers and depending on the amount of data and its dimensionality, the runtime for a single optimization is in the order of seconds and thus allows for multiple restarts. We have found 10 to 15 restarts sufficient for the algorithm to arrive at the same optimum several times.

6.2 Validation

In this section, we describe the simulations designed to assess the performance of cSPoC as well as conventional approaches in a broad range of realistic scenarios. Moreover, we present three real-data studies well-suited to demonstrate cSPoC's effectiveness in recovering brain oscillations in a variety of practical applications.

6.2.1 Simulations

The data generation processes and the evaluation metrics are identical to what was described in chapter 3.2. The methods are compared in terms of *accuracy* and *interpretability*, where accuracy describes how well the functional activation of the target source components is extracted and interpretability reflects the similarity of the estimated activation patterns to those of the true target components.

For each simulation run, we created the datasets $\mathbf{x}$ and $\acute{\mathbf{x}}$ according to the generative model Eq. (2.1), where the details of the data generation have been described in chapter 3.2. The number of simulated channels was $N_x = N_{\acute{x}} = 25$, and a total of $K_x = K_{\acute{x}} = 31$ sources was simulated for each data set. One source from $\mathbf{x}$ was functionally coupled with one source from $\acute{\mathbf{x}}$, such that the envelope of this *target source* in $\mathbf{x}$ was correlated with the envelope of the respective other *target source* in $\acute{\mathbf{x}}$. The time courses (and envelopes) of all other sources were completely independent to each other and to the target sources. These independent sources are referred to as background sources.

Consistent with the simulations of the preceding chapters, we assessed the performance of the compared methods with respect to (i) varying signal-to-noise ratios (SNR) between the target source and the background sources, (ii) varying amounts of training data available, and (iii) to the true degree of coupling between the target source components from $\mathbf{x}$ and $\acute{\mathbf{x}}$, i.e. to the true correlation between $\mathbf{x}$ source envelopes and $\acute{\mathbf{x}}$ source envelopes. This last aspect pertains to how the algorithms perform as the underlying assumption (functional coupling between modalities) are no longer valid.

The amount of training and test data was three and two minutes, respectively, for each of the simulation runs in which either the SNR or the true underlying coupling was varied. In runs in which the amount of training was varied, the SNR was set to -10 dB. For each parameter setting the simulations were repeated 150 times with newly generated data.

Data analysis

We benchmarked cSPoC against other unsupervised methods, namely CCA, PowerCCA (Ramírez et al., 2013), and ICA, as well as one supervised method, namely ordinary-least-squares (OLS).

OLS was applied to each modality separately and, being a supervized method, it received the time-courses of the true target sources as additional training data. Thus, OLS again serves as an empirical upper bound in the performance comparison.

ICA was applied to the training data of each modality separately. Thereafter, the envelopes of the $\mathbf{x}$ ICA components obtained on the training data were computed and correlated with the envelopes of the $\mathbf{y}$ ICA components. The pair that yielded highest correlation on the training data was then assessed on the test data.

CCA was applied as outlined in figure 6.1, i. e., envelopes were computed for each simulated channel, and the envelope time courses were used as input for CCA. All other methods act on the band-limited time-domain data, and envelopes were computed after performing the backward projection, i. e., on the extracted sources.

PowerCCA is a recent method that also optimizes spatial filters for two data sets such that the envelopes of the projected signals exhibit maximal correlation. We discuss the relation between CCA, PowerCCA and cSPoC in the Discussion section of this chapter.

Results

The results of the simulations are shown in figure 6.2, figure 6.3, and figure 6.4. Since the optimization problem was symmetrical with respect to the datasets $\mathbf{x}$ and $\acute{\mathbf{x}}$, the figures show only to what extend the methods were able to recover the $\mathbf{x}$ target source. In addition, the figures show the estimated coupling between modalities, i.e. the envelope correlation obtained on the estimated $\mathbf{x}$ and $\acute{\mathbf{x}}$ source components.

Figure 6.2 shows the performance of the compared methods as a function of SNR for two different settings of true underlying coupling (correlation 1 and correlation 0.7). Both cSPoC and PowerCCA outperform CCA and ICA over wide ranges of SNR in all three performance metrics. Comparing cSPoC and PowerCCA, it is cSPoC who is consistently better in the recovery of the target source. Similar to what we have seen before, CCA is unable to accurately extract the amplitude modulations because its application here is not in line with the generative model.

A very similar picture can be seen in figures 6.3 and 6.4, which depict the performance as a function of true underlying coupling or amount of training data available, respectively. cSPoC and PowerCCA outperform the other methods because their assumptions best correspond to the (realistic) generative model of the data and they use a maximal amount of information available, which is in this case information provided by the respective other modality. In the direct comparison between cSPoC and PowerCCA, it is cSPoC that consistently outperforms PowerCCA. Possible reasons for this will be discussed at the end of the chapter.

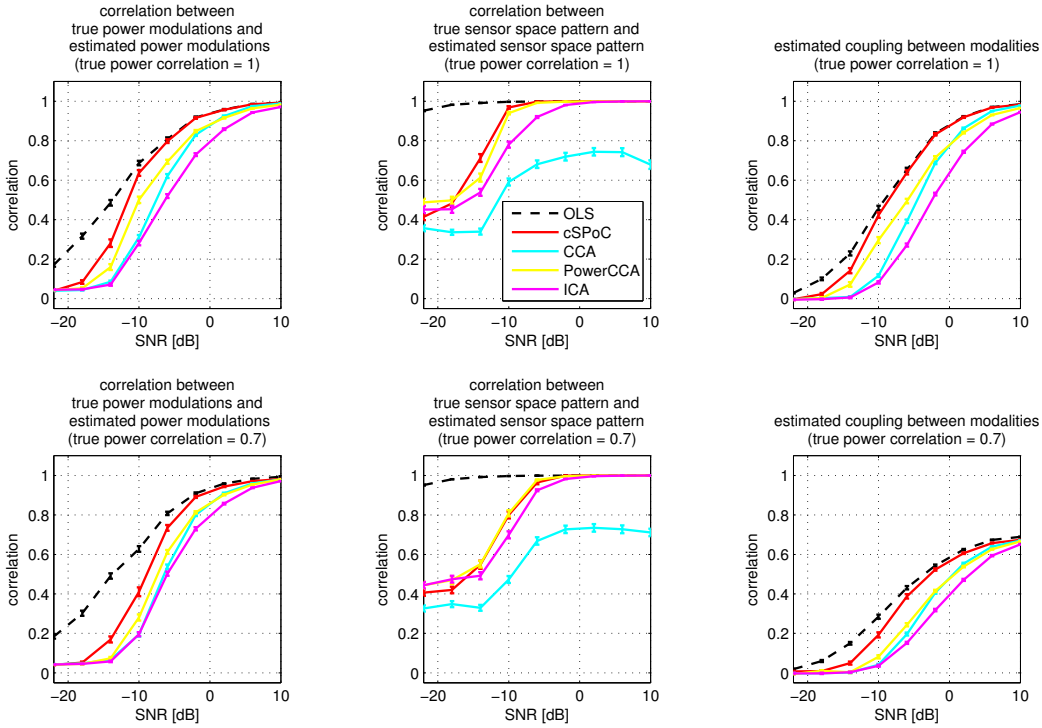


Figure 6.2: Simulation results. Here performance in terms of accuracy and interpretability is shown as a function of signal-to-noise ratio. The right-most plot shows the bandpower correlation between the estimated target sources. In the top row of the plots, the bandpower correlation between the true target sources was 1, while in the simulations depicted in the bottom row, that correlation was 0.7.

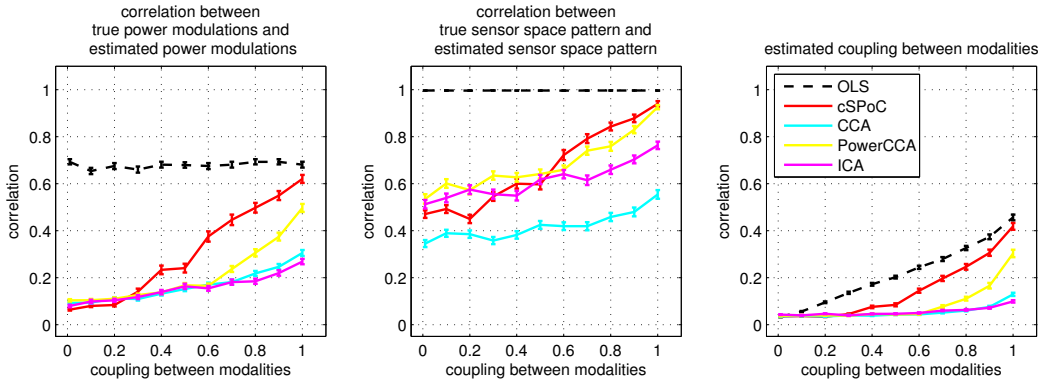


Figure 6.3: Simulation results. Here performance in terms of accuracy and interpretability is shown as a function of true underlying coupling. The SNR was -10 dB in these simulations and 600 epochs (corresponding to 5 minutes of data) were used to train the methods.

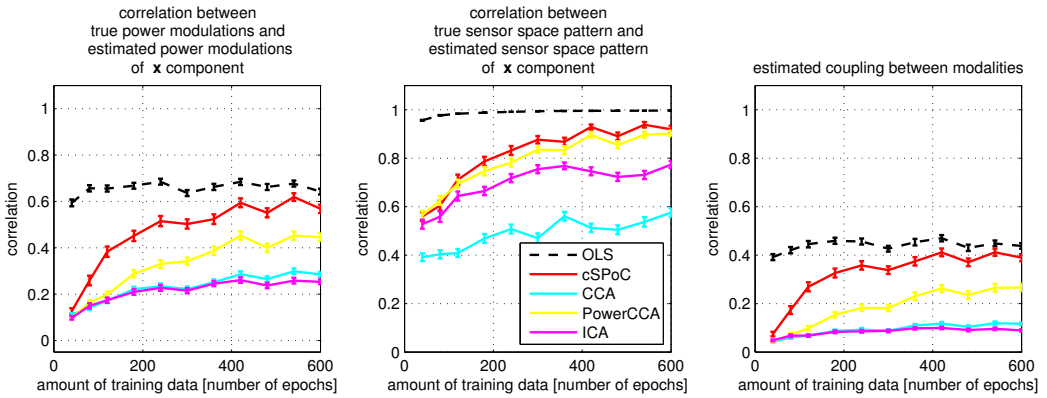


Figure 6.4: Simulation results. Here performance in terms of accuracy and interpretability is shown as a function of amount of training data available. Training data is measured in number of epochs, each of which was 500 ms long in these simulations. Epochs did not overlap and the SNR was set to -10 dB.

6.2.2 Real data examples

Application example I: Inter-subject envelope correlations during auditory stimulation

As an initial real-world example, we here consider the so-called *Hyperscanning* setting, where envelope coupling of brain oscillations arises, because multiple subjects are stimulated in the same way. In the absence of any other information, this coupling can be utilized by cSPoC to extract the underlying oscillatory processes. Importantly, this can be done even if neither the stimulus sequence nor the actual relationship between stimulus and response is known.

In order to show this using real data, we here re-analyze the same data that was used to validate the SPoC algorithms in chapter 4.2.2.

The SPoC method described in chapter 4 is a supervised method relying (i) on knowledge of the stimulation intensity (in this case loudness) and (ii) on the assumption of a *linear* relationship between stimulus intensity and the envelope of the oscillatory brain response. While both of these prerequisites are fulfilled for the considered SSAEP data, this might not be the case in general. Here we show that competitive performance can be obtained using cSPoC in a Hyperscanning setting, i. e., when multiple datasets with identical stimulation are available. Notably, cSPoC is an unsupervised method neither requiring any information about the stimulation, nor assuming a particular relationship between stimulus and brain response.

Preprocessing We applied spatio-spectral decomposition (SSD, Nikulin et al., 2011) to perform dimensionality reduction by projecting the data onto the space spanned by the 10 components showing largest spectral peaks in the frequency band of interest, i. e., around

40 Hz. The SSD-projected data were then band-pass filtered using a pass-band of 39 to 41 Hz.

The actual loudness modulations were not identical for all subjects in the original experiment. However, all modulation functions had the same uniform histogram over the presented loudness interval, which means that all subjects were exposed to the same intensity levels for the same number of times, only in a different order. Thus it is possible to create a *virtual* same-stimulus scenario from this data. In order to do so, the EEG data as well as the corresponding loudness modulation were segmented into consecutive non-overlapping epochs of 1 second length. For each pair of subjects S_i and S_j, the epochs of subject S_j were re-ordered such that the re-ordered loudness function of subject S_j became aligned with true loudness modulation function of subject S_i.

Application of cSPoC and SPoC Since we did not expect between-subjects EEG envelope correlations *within* epochs but only *across* epochs, we applied the cSPoC version that optimizes correlations across epochs as introduced in section 6.1.2.

We applied cSPoC to all pairwise combinations of the data of the seven subjects. For each subject pair, only the first, i.e. highest envelope-correlating, source pair was extracted. SPoC, on the other hand, was applied to the data of each subject separately, using the corresponding loudness modulation as the target function, as done in Dähne et al. (2014d). Due to this knowledge, SPoC has a crucial advantage over cSPoC.

The performance of both algorithms is measured in terms of the correlation of the envelopes of the extracted sources with the corresponding target variables, as well as by computing all possible pairwise correlations between the source envelopes obtained from different subjects. In order to exclude overfitting effects, we report average correlations obtained in a 10-fold cross-validation scheme.

Results Figure 6.5 shows cross-validated correlation coefficients obtained by SPoC and cSPoC in the re-analysis of the auditory steady-state data. The subjects are ordered by the correlations between stimulus intensity and envelope as obtained with SPoC, from highest to lowest. Plot **A** in the figure shows the correlations between the SPoC/cSPoC component envelope and the loudness modulation for each subject (mean over cross-validation folds), as well as the average correlation across subjects, which was 0.64 ± 0.13 (mean $\pm$ SEM) for SPoC and 0.58 ± 0.17 for cSPoC. Plot **B** shows the average over subject pairs, which was 0.4 ± 0.11 for SPoC and 0.35 ± 0.16 for cSPoC. Plots **C** and **D** show the pairwise envelope correlations for SPoC and cSPoC, respectively (mean over cross-validation folds).

Although having no access to the loudness modulation of the stimulus, cSPoC delivers correlations that are on par with the correlations obtained by SPoC. The gap between the correlations obtained by the two methods widens as the correlation between the envelope of the extracted signal and the target function becomes weaker. This may be caused by lower SNR or lower correlation between the SSAEP component and the loudness modulation. In either case the effect is stronger on the performance of cSPoC than on SPoC. The effect is consistent with observations from the results of the simulations.

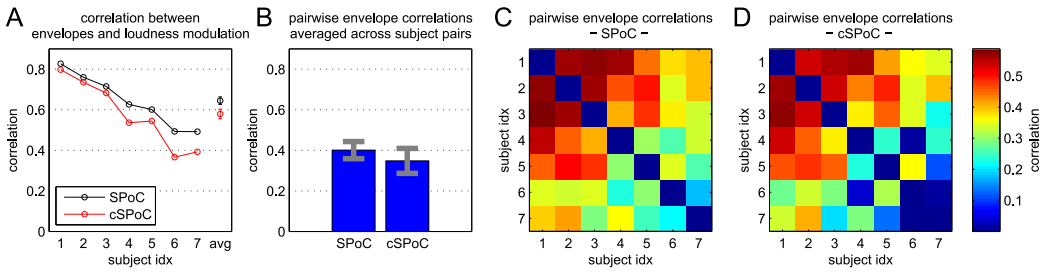


Figure 6.5: Results of inter-subject envelope correlation in an auditory steady-state paradigm: correlations. The unsupervised cSPoC method was applied to pairs of subjects, without using the actual loudness modulation. The supervised SPoC method was applied to each subject individually, using it's corresponding loudness modulation as a target signal. (**A**) Cross-validated correlations between envelopes of sources extracted by SPoC/cSPoC and loudness modulation (stimulus intensity) for each subject, as well as average across subjects. Errorbars indicate SEM. (**B**) Pairwise envelope correlations for all subjects as obtained by SPoC and cSPoC, averaged across subject pairs. Errorbars indicate SEM. (**C**,**D**) Pairwise envelope correlations for all subjects as obtained by SPoC and cSPoC, respectively.

The spatial activation patterns of the SSAEP components extracted by each method from each subject are displayed in Figure 6.6. The polarity of the SPoC/cSPoC patterns is arbitrary and thus was set such that highest magnitude value is positive. We observe almost no differences in the spatial patterns obtained for individual subjects by cSPoC and SPoC. The spatial patterns show the characteristic maxima (minima) at mid-frontal electrode positions and minima (maxima) at posterior electrode positions, with the sign depending on the phase of the stimulus. Such patterns have previously been reported in event-related potential (ERP) analysis and source modeling of SSAEP responses by Herdman et al. (2002).

Summarizing, the results obtained in this analyses show that cSPoC is able to reliably extract source activity whose envelope dynamics are stimulus induced and consistent across subjects, *without knowledge about the stimulus.*

Application example II: Cross-frequency coupling of alpha and beta oscillations during rest

Cross- and within frequency interactions are hypothesized to be an important mechanism for integration of neuronal processes distributed across different brain areas and frequency bands. In this experiment, we focused on envelope interactions between and within alpha and beta oscillations during rest.

Experimental paradigm, data acquisition and preprocessing Seven subjects participated in the study (2 females). EEG recordings were performed at rest condition (15 min) with subjects seated comfortably with their eyes being open. For dimensionality reduction the

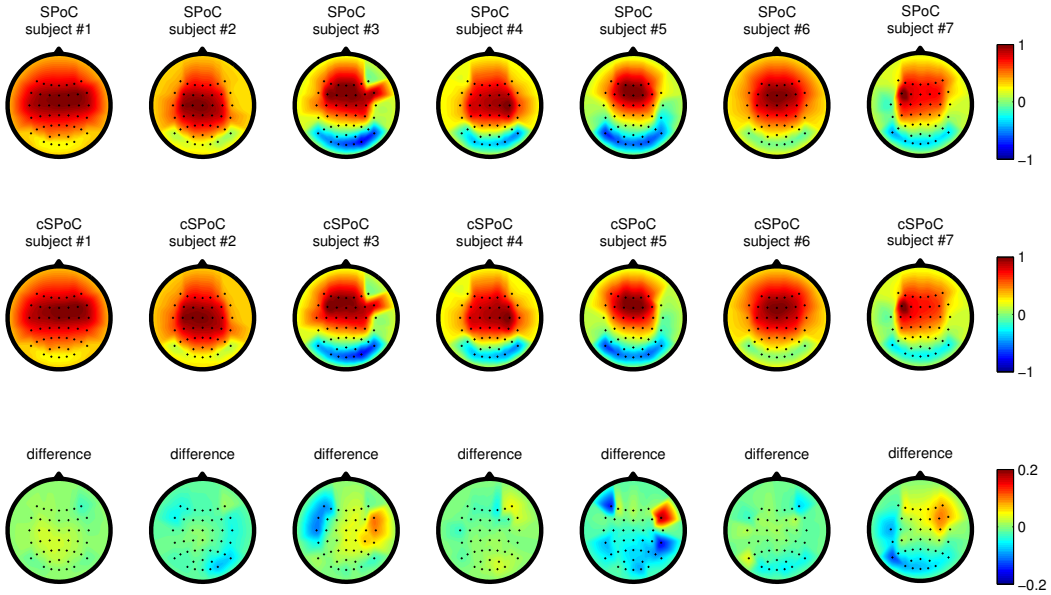


Figure 6.6: Results of inter subject envelope correlation in an auditory steady-state paradigm: spatial patterns. For each subject, the top row shows spatial patterns obtained by SPoC, while the second row shows patterns obtained by cSPoC. The third row shows the difference between the patterns of the first and second row. Note that the scaling of the patterns (i.e. magnitude and polarity) is arbitrary in both algorithms. Here the scaling was adjusted such that the largest magnitude was 1. Color scale is the same in rows one and two, but note the different color scale in row three.

data were projected to two separate 10 dimensional subspaces: the first SSD subspace captured oscillatory activity in the alpha range (9 to 12 Hz) and the second SSD subspace captured oscillatory activity in the beta range (19 to 23 Hz).

The SSD-projected data were bandpass filtered separately in the alpha range (9 to 12 Hz) and in the beta range (19 to 23 Hz), yielding two 15 minutes long narrow-band data sets per subject. For each subject, cSPoC received the alpha and beta band-passed data as input and five spatial filter sets were optimized per subject. Envelopes of the spatially filtered signals were averaged within consecutive, non-overlapping windows of one second length and the resulting envelope time-courses were correlated. The entire analysis was carried out in a cross-validation loop to avoid overfitting.

Results Figure 6.7 shows the cross-validated envelope correlations between the first five cSPoC component pairs, i. e., the correlation between the envelopes of the alpha and beta band source found by cSPoC. In five out of seven subjects, it can be observed that the first two component pairs deliver notably higher correlations between alpha and beta compared to the following pairs. For the remaining two subjects (subject 3 and 4), this is true for the

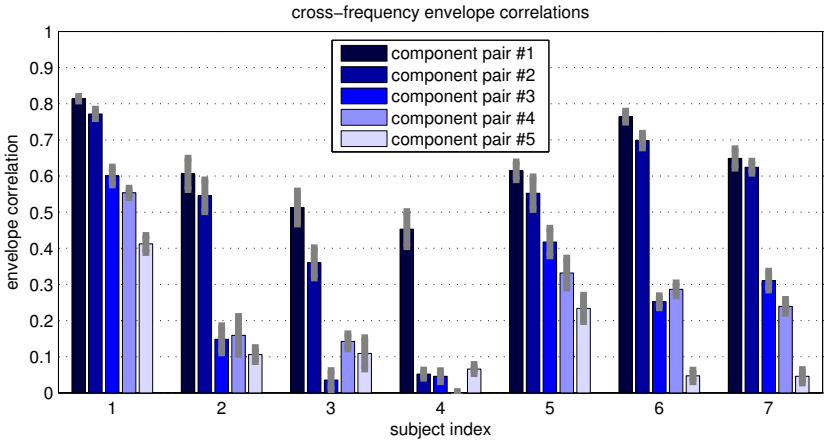


Figure 6.7: Results of resting state alpha/beta analysis: cross-frequency envelope correlations. Cross-validated correlations between the envelopes of alpha and beta oscillations of the first five cSPoC component pairs for each subject. See Figure 6.8 for corresponding spatial patterns of selected subjects. Errorbars indicate SEM over cross-validation folds.

first component pair. Across subjects, the envelope correlations of the first component pair are in the range between 0.4 and 0.8.

The spatial activation patterns of three representative subjects are plotted in Figure 6.8. The figure reveals that there is a remarkable similarity of spatial patterns of the alpha and beta band sources within each source pair found by cSPoC, where the similarity is strongest for the component pairs that exhibit highest correlation. This similarity between the spatial patterns indicates that the corresponding neural generators of oscillatory alpha and beta activity are in very close proximity to each other in the brain, or perhaps even occupy the same cortical space.

In six out of seven subjects, the spatial patterns indicate sensorymotor-related areas as the origins of brain sources with the most strongly coupled envelope dynamics (see subjects 1 and 7 in the figure for example). However, in subject 5 the spatial patterns suggest sources in parietal or occipital areas to exhibit the largest envelope correlations.

Summarizing, the analysis of resting-state data using cSPoC revealed pairs of alpha- and beta band oscillations with large envelope correlations and highly similar physiologically plausible spatial activation patterns. The extracted signals serve as an excellent basis for further investigating the relationship between alpha and beta oscillations, for example by assessing phase synchrony or other measures.

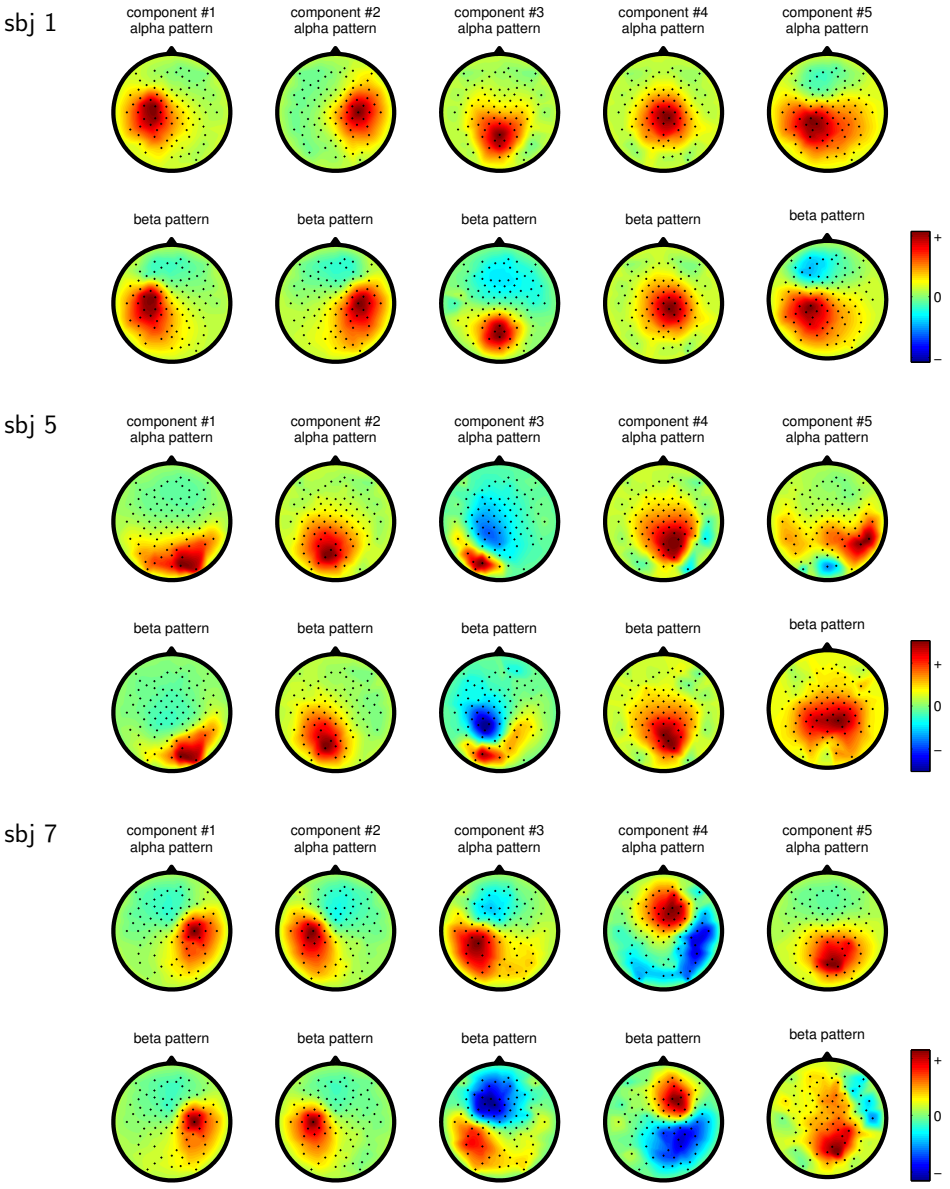


Figure 6.8: Results of resting state alpha/beta analysis: spatial patterns. Spatial activation patterns of the five cSPoC components pairs with best correlating alpha/beta envelopes. For each of the three subjects presented, the top row shows the spatial patterns of components in the alpha band while the bottom row shows the respective best correlating component in the beta band. The envelope correlations between the alpha/beta pairs can be found in figure 6.7.

Application example III: Multimodal integration of simultaneously measured EEG and MEG

In the final application example of cSPoC we demonstrate the extraction of meaningful components from simultaneously measured EEG and MEG recordings during rest.

Experimental paradigm, data acquisition and preprocessing Here we analyze the data of a single subject who was seated comfortably in an MEG scanner while wearing an EEG setup at the same time. Both EEG and MEG was recorded simultaneously for a duration of 10 minutes, during which the subject was asked to relax with open eyes. EEG data consisted of recordings obtained from 55 channels arranged on the scalp according to the international 10-20 system. MEG signals were recorded with 306-channel Vectorview MEG system (Elekta Oy, Helsinki, Finland). Motion compensation and artifact removal of MEG was done by the tSSS method (Taulu and Simola, 2006). The MEG system has three types of sensors that each measure a different aspect of the local magnetic field. These sensor-triplets are placed at 102 locations across the scalp, which together yields 306 MEG channels. Here we only used the data from one sensor per triplet and location, yielding 102 channel MEG data. The sensor used is the so-called magnetometer, which measures magnetic field strength perpendicular to the surface of the scalp.

The aim of this analysis is to extract source components that are observable with both measurement modalities and should therefor have highly correlated envelopes when extracted from the respective modality. Using the steps described below, the following frequency bands were analyzed separately: 5 Hz to 8 Hz, 8 Hz to 13 Hz, 12 Hz to 19 Hz, 16 Hz to 25 Hz, 20 Hz to 32 Hz, and 24 Hz to 38 Hz. Respectively, these bands roughly correspond to what is referred to in the EEG/MEG literature as the theta-, alpha-, lower beta-, middle beta-, high beta-, and low gamma band.

MEG and EEG data were separately subjected to SSD preprocessing, retaining 20 SSD components per modality. Thereafter, the continuous sensor space data was segmented into one second long and non-overlapping epochs, yielding a total of 600 epochs. Epochs that exhibited excessive amounts of variance (averaged across channels) were removed. cSPoC was then employed to extract EEG/MEG component pairs that exhibit maximal envelope correlations. Note that – in order to suppress noise – cSPoC component envelopes were first averaged within epochs and then correlations were computed using the epoch-wise averaged envelopes.

In order to asses the significance of the obtained envelope correlations, a permutation test approach was used. To this end, surrogate data was created by simply shuffling the order of EEG epochs, while keeping the order of MEG epochs intact. Importantly, the shuffling procedure that was used preserved the auto-correlation of global envelope dynamics (Theiler et al., 1992). Thereby, all aspects of the data were retained, except for temporal alignment between modalities. The cSPoC analysis was repeated 200 times using surrogate data in order to generate a distribution of envelope correlation values to which the correlations obtained on the un-permuted data were compared.

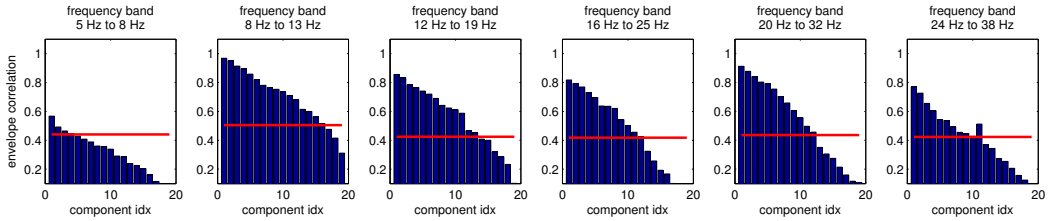


Figure 6.9: Results of EEG/MEG analysis. Each plot shows the envelope correlations between cSPoC EEG and MEG components obtained for separate frequency bands, where the band limits are given in the title of each plot. The red line indicates the 99.9% percentile of an envelope correlation distribution that was obtained via band-specific permutation analysis.

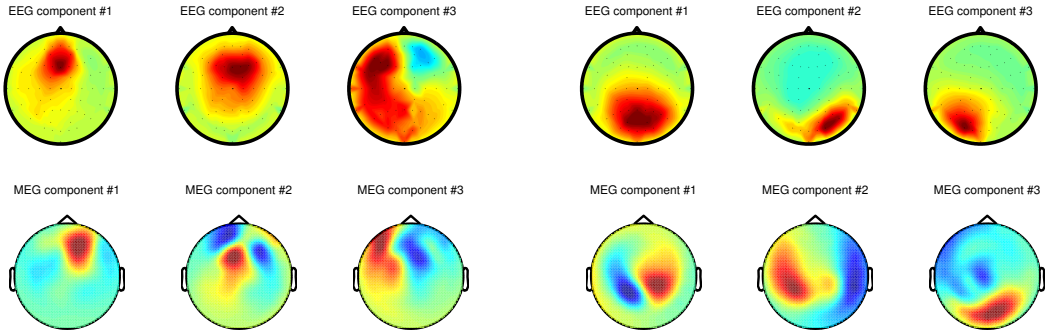
Results Figure 6.9 shows the obtained envelope correlations separately for each frequency band. In each plot, we also show the 99.9% percentile of the correlation distribution obtained from permutation analysis. The largest envelope correlations are obtained for the alpha band (up to 0.96) while the theta band shows much lower correlations between EEG and MEG power (here the maximum is 0.57). Also, the number of significantly correlated component pairs is lowest for theta and highest for alpha. The next-strongest coupling between modalities can be observed in the beta range.

Figure 6.9 shows the first three cSPoC component pairs from each of the six frequency bands. The strongest envelope correlations among the theta components is exhibited by components with frontal topographies. The cSPoC alpha components, on the other hand, are consistent with neural sources in parietal and occipital regions of the brain. The spatial patterns of the beta components suggest neural sources near or in sensorimotor areas of the brain.

All of these brain areas are part of what has become known as the default-mode network (see Raichle et al., 2001; Raichle and Snyder, 2007, for example). Thus our analysis has delivered a set of components from a number of frequency bands that allow to study the electrophysiological correlates of concerted brain activity during rest.

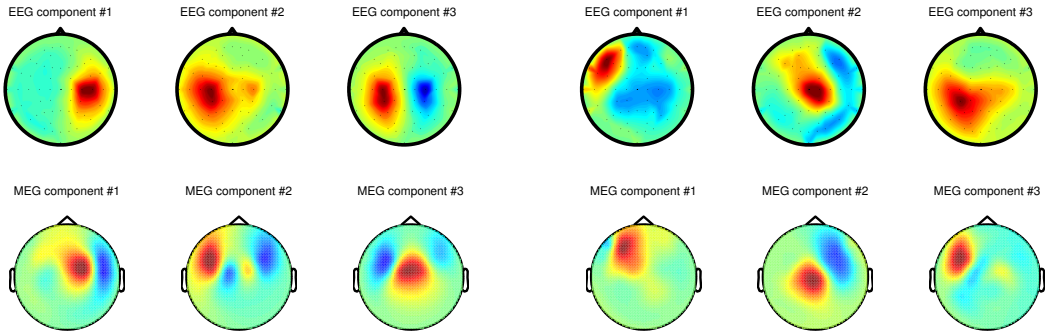
6.3 Discussion

In this chapter, we have presented cSPoC, a novel unsupervised source separation approach for finding oscillatory sources with envelope correlations. The method was benchmarked using simulations as well as real EEG data. Its performance, which is competitive to supervised approaches, makes cSPoC the method of choice in scenarios in which supervised methods are not applicable, such as Hyperscanning settings. It thus provides a versatile addition to other multivariate analysis tools for cross-frequency coupling such as cross-frequency decomposition (CFD, Nikulin et al., 2012) for phase coupling, as well as tools for extracting brain activity from electrophysiological recordings based on other types of dependencies (Nolte et al., 2006, 2008; Dyrholm et al., 2007; Gómez-Herrero et al., 2008; Haufe et al., 2010).



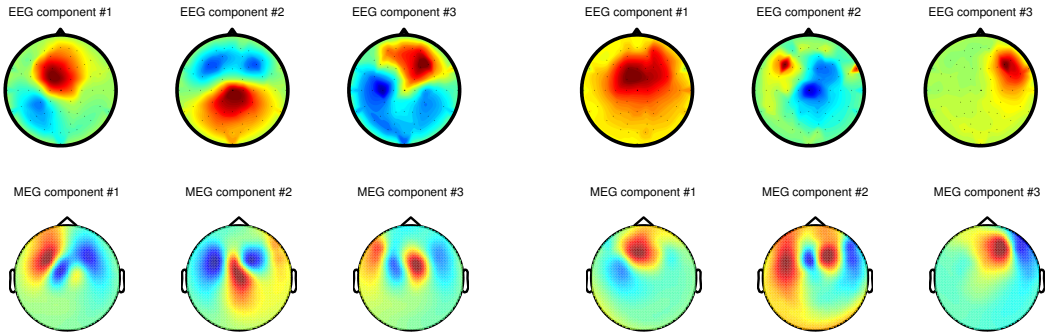
(a) 5 Hz to 8 Hz

(b) 8 Hz to 13 Hz



(c) 12 Hz to 19 Hz

(d) 16 Hz to 25 Hz



(e) 20 Hz to 32 Hz

(f) 24 Hz to 38 Hz

Figure 6.10: Results of EEG/MEG analysis. Each of the sub-figures (a to f) shows the spatial activation patterns of the first three cSPoC component pairs obtained for a specific frequency band. The band limits are given in the captions of the sub-figures. The corresponding envelope correlations are shown in figure 6.9.

Being a multivariate technique, cSPoC integrates information from all recording channels and thereby achieves much higher SNRs compared to single channel analysis, which leads to stronger effects. By projecting the data into the cSPoC source space, the dimensionality is reduced and thus the problem of channel-wise comparison/correlations (and multiple testing) is avoided. In our analysis examples, we demonstrated the robustness of obtained envelope correlations by means of cross-validation and permutation tests. While statistical approaches have been developed to assess functional connectivity after projection into a source space (e.g. Carbonell et al., 2009), these approaches typically test the connectivity of source time-courses, rather than connectivity of their envelopes. Thus, it is desirable to extend such statistical models in order to avoid potentially time-consuming permutation analysis.

Related methods Canonical Source Power Analysis is conceptually and structurally related to a number of existing methods. Among those are two recent approaches also aiming to extract brain oscillations with power dependencies, PowerCCA by Ramírez et al. (2013) and non-linear CCA by Campi et al. (2013). We included PowerCCA in the simulations, where it was outperformed by cSPoC. We believe the reason for that to be the different approaches for optimizing envelope correlations employed by cSPoC and PowerCCA. PowerCCA optimizes over huge $N^2_{\mathbf{x}/\mathbf{y}}$-dimensional parameters matrices, where $N_{\mathbf{x}}$ and $N_{\mathbf{y}}$ are the number of recording channels for data sets $\mathbf{x}$ and $\mathbf{y}$. The actual extraction filters are obtained post-hoc as the first eigenvectors of these matrices. In contrast, cSPoC optimizes over filter vectors directly and has thus only $N_{\mathbf{x}} + N_{\mathbf{y}}$ parameters to tune. For this reason, cSPoC may be less prone to overfitting than PowerCCA. Non-linear CCA by Campi et al. (2013), which is related to an earlier approach by Gutmann and Hyvärinen (2011), essentially computes linear CCA in the space of all binomials of the original input space, with a dimensionality constraint on the weight vectors. While their method achieves good results on combined data of 2 x 5 subjects, it failed to do so on single subject-pair level.

There is also an interesting relation between cSPoC and algorithms that optimize higher order (cross-) moments of the projections. Among these algorithms are a specific class of independent component analysis (ICA) algorithms (Comon, 1994; Cardoso and Souloumiac, 1993b; Hyvärinen and Oja, 1997), as well as minimum overlap component analysis (MOCA) (Nolte et al., 2009). These algorithms may arrive at similar solutions using an entirely different set of assumptions and constraints. Yet, they are not explicitly designed to optimize envelope correlations of projections of multivariate oscillatory signals.

Clinical perspective cSPoC can also be used in clinical studies. For instance recent studies in patients with Parkinson disease revealed an existence of phase-synchronization in beta frequency range between subthalamic nucleus (STN) and cortex (Klostermann et al., 2007; Litvak et al., 2011). Here an interesting question would be how the envelope of alpha oscillations (reflecting cortical excitability) might affect local neuronal dynamics in STN, the latter being expressed in the amplitude of beta oscillations. These alpha versus beta envelope interactions can be studied with cSPoC.

Conclusion We presented a novel method, cSPoC, which is able to identify brain oscillations with correlated envelopes from high-dimensional and noisy electrophysiological recordings. The method performs excellent in simulations and showed very good results in three distinctively different, yet highly relevant, analysis scenarios involving real EEG data. We believe that cSPoC will be a valuable tool in the quest for understanding the mechanisms and functions of power-to-power coupling in electrophysiological recordings.

Chapter 7

Discussion and conclusions

In this final chapter we shall now give a more broad discussion of the presented methods and ideas. The novel analysis methods derived in this thesis — SPoC, mSPoC, and cSPoC — will be collectively referred to as the SPoC framework. Next to discussing some general strength and weaknesses, a particular focus of this chapter is to place the SPoC framework within a number of taxonomies, each constructed from a different perspective. Thereby we provide an intuition of where and how the methods of the SPoC framework fit into the 'big picture' of neuroimaging analysis methods.

When to use which method?

Given the multitude of analysis approaches available, the most relevant question for the practitioner is of course which method to choose for a specific research question. Unfortunately there is no unique answer to this, because it depends on (i) the preferred analysis scenario (e.g. late fusion vs early fusion), (ii) the assumptions being made about the data (e.g. what type of coupling between modalities), (iii) the features of interest in the analysis (e.g. spectral features or time-domain features), (iv) what additional information is available (e.g. condition/class labels), and other aspects. However, in order to narrow down the possible choices we provide a categorization of the methods discussed in this thesis in table 7.1. The categorization scheme of table 7.1 — though by no means exhaustive — is designed to best reflect the perspective of a neuroscience researcher who is searching for an appropriate analysis approach.

When will a method work and, more importantly, when will it not work?

This question relates to the underlying assumptions and to the limits of an analysis method. As for the SPoC framework, the two most important assumptions are these: (i) the data can in fact be decomposed into components, i.e. the data is sufficiently described by a factor model as outlined in chapter 2.2, Eq. (2.1) and (ii) there indeed is a functional relation between the amplitude modulations of a component and a given other signal, i.e. the assumed co-modulation does in-fact exist. Let us denote these assumptions by *linear decomposability assumption* and *co-modulation assumption*, respectively. It has to be clearly stated that if these two assumptions are not met, the SPoC framework cannot be expected to deliver accurate and meaningful results, because in such a situation it would simply

late fusion scenarios

features of interest	amplitude modulations	ERPs or hemodynamics
suggested methods	SPoC, CSP, SSD, ICA, PCA	Regression, GLM, LDA, ICA, PCA

early fusion scenarios

features of interest	amplitude modulations and hemodynamics	ERPs and hemodynamics
suggested methods	mSPoC	CCA/PLS, jICA

Table 7.1: Suggestions for when to choose which of the methods discussed in this thesis.

not be an adequate model for the data. In case of the decomposability assumption not being true, this negative result is essentially the same for all of the other decomposition methods.

However, the decomposability assumption has been the basis for meaningful results obtained from macroscopic measurements of brain activity many times over and is at the heart of a large number of analysis techniques for neuroimaging data. See the reviews by Hämäläinen et al. (1993); Baillet et al. (2001); Nunez and Srinivasan (2005) and the work cited therein for considerations in favor of the decomposability assumption. However, we would also like to point out recent work by Hindriks et al. (2014) which challenges an important prerequisite of the decomposability assumption, namely the existence of stationary (i.e. fixed) spatial activation patterns.

The co-modulation assumption, i.e. that of an existing functional relation between either different measurement modalities or datasets of different subjects, is a matter of ongoing research. In our simulations we have demonstrated the behavior and performance of our methods as a function of true underlying co-modulation strength. We have seen that the methods no longer recover the desired target components once the assumed co-modulation is too weak. The minimum level of co-modulation necessary for extraction of target components, however, depends on other factors, such as the amount of data and the signal-to-noise ratio. On real data, it is of course unclear if and to what extend a co-modulation exists. Here, however, the SPoC framework can contribute by providing a testbed for the co-modulation assumption. By first performing a (m/c)SPoC analysis on the data in its correct temporal order and then repeating the analysis on temporally shuffled version of the same data, a distribution of correlation values can be obtained that represents the null-hypothesis of no co-modulation being present. This allows for direct assessment of the co-modulation assumption and thereby provides researchers with tool to study this questions from many possible angles. In addition, the bonus of linear decomposition methods is that — in case of confirmed co-modulation — the involved source components can be inspected and analysed with respect to anatomical origin of the coupled signals from the involved modalities.

What about fusing more than two imaging modalities?

Recall that the SPoC algorithm required the presence of a univariate target signal, which could, among others, encode label information. Furthermore, the mSPoC and cSPoC algorithms were defined for the fusion of two datasets. It turns out that due to the conceptual similarity between the three members of the SPoC framework, it is in principle possible to merge them into a single algorithm. While the measures of co-modulation that were employed here (covariance and correlation) are defined between two variables only, we can make use of a trick often used to generalize algorithms to sets of more than two variables. This simple trick consists of maximizing the sum of pair-wise co-modulation measures and was proposed in the context of CCA, for example, in Kettenring (1971). Based on this simple idea, a single objective function can be derived that contains the entire SPoC framework, as well as regression, LDA, CSP, and CCA as special cases. The derivation of this objective function contains only elements and ideas that have already been presented earlier in this thesis as well as in the literature outside this thesis. Thus, the details of the objective function can be found in Appendix B. Importantly however, this approach represents a step towards early fusion of more than two functional imaging modalities. Corresponding datasets exist, such as for example a dataset published in Wakeman and Henson (2015), which contains concurrent (i) EEG and (ii) MEG measurements of multiple subjects undergoing a visual detection paradigm as well as (iii) fMRI measurements of the same subjects executing the same experimental paradigm in an MRI scanner. The concurrent EEG and MEG data can be functionally aligned to the corresponding fMRI data for each subject and thereby represent a tri-modal functional imaging dataset.

Using more than just functional information: Incorporating prior anatomical knowledge.

The detection of functional coupling lies at the heart of the SPoC framework (recall the co-modulation assumption from before). In this chapter we have outlined how the SPoC framework can be used to probe for the existence of the proposed functional coupling and throughout this thesis we have used *anatomical links* between modalities only to validate the meaningfulness of the extracted components. That is to say, we have judged the success of component extraction (that was based on assumed functional relations) by whether the spatial patterns of extracted components also made sense in anatomical relation to each other. While reserving available anatomical information for validation of results may be a cautious way to proceed, it is also a potentially sub-optimal one, because information that is potentially useful in the extraction of components is withheld from the algorithm. Such anatomical information is often available in the case of concurrent EEG and fMRI measurements and the link between brain anatomy, as measured during a structural MRI scan, and expected topologies of components in EEG/MEG sensor space is given by the so-called leadfield matrix.

In recent efforts, we have begun to include prior anatomical knowledge into the mSPoC algorithm in order to overcome a potential weakness due to the focus on functional links between modalities only. Specifically, we propose to extend mSPoC (when working with EEG and fMRI) by biasing the spatial filter of the EEG component towards brain locations

that exhibit large magnitudes in the spatial activation pattern of the corresponding fMRI component. Details of these considerations as well as simulations that support our ideas have been published in Hansen et al. (2015). The proposed extension to the mSPoC algorithm shows greatest performance improvements when the signal-to-noise ratio is low and only small amounts of training data are available.

Going beyond correlative coupling: Incorporating measures of causality

An approach to studying the *causal effects* of neural oscillations is the following: identification of causal relationships based on temporal precedence as revealed by a concept called 'Granger causality' (Granger, 1969). Granger causality is a standard statistical method from the field of econometrics and has been applied in neuroscience to infer functional brain connectivity (e.g. Roebroeck et al., 2005; Astolfi et al., 2007; Nolte et al., 2008; Bressler and Seth, 2011). Using the previously introduced notation in which ϕ represents neural amplitude modulations z denotes a variable of interest, ϕ is said to *Granger cause* z if ϕ helps to predict the future z above what is predicted by the past of z alone. In the context of EEG recordings, for example, an actively researched question in the field of Brain-Computer Interfaces (BCI) is whether (and how) oscillatory sources influence the control performance of a user during a BCI experiment (Grosse-Wentrup et al., 2011; Dähne et al., 2011a; Maeder et al., 2012).

In Winkler et al. (2015), we investigated which methods are bests suited to reveal a Granger causal effect from neural oscillations to a given external target variable. To this end, we compared channel-wise Granger causality testing with three source separation methods: ICA, SPoC, and a novel analysis method which extracts a source whose bandpower maximally Granger causes the target variable. This last method is called *Granger Causal Power Analysis* (GrangerCPA) and it was derived based on concepts developed in this thesis. In Winkler et al. (2015) GrangerCPA is found to outperform ICA and SPoC in the extraction of source components whose amplitude modulations are causally related to a given target signal. The comparison was carried out in simulations and on two real EEG data sets.

How does the SPoC framework fit into the 'big picture' of decomposition methods?

Finally, we make two attempts to place the SPoC framework into categorization schemes that show how the methods of the SPoC framework relate to other decomposition methods that are available. Note that there is one thing all of the methods compared here have in common: they optimize weight vectors that are applied to the data in sensor space (or a linearly preprocessed version thereof), i.e. we have $\mathbf{w}^\top\mathbf{x}$, where $\mathbf{x}$ is *not* non-linearly preprocessed.

		type of additional information		
		binary labels	univariate target signal	other multivariate signal
signal aspect of interest	$\mathbf{w}^\top\mathbf{x}$	LDA	Regression	CCA, PLS
	Phase $(\mathbf{w}^\top\mathbf{x})$	CSIP	?	MaxImC
	Ampl $(\mathbf{w}^\top\mathbf{x})$	CSP	SPoC, Grangercausality-CPA	mSPoC, cSPoC, PowerCCA

Table 7.2: A categorization of decomposition methods according to what type of additional information is used for the decomposition of a dataset and which aspect of the projected signal is of interest for the decomposition.

The first categorization scheme we chose is depicted in table 7.2 and collects a set of supervised and semi-supervised decomposition methods and categorizes them based on two properties: (i) the signal aspect of interest and (ii) the type of additional information that is provided to the algorithm. *Signal aspect of interest* here refers to what property of the projected signal $\mathbf{w}^\top\mathbf{x}$ is relevant for the objective function of the method. The three categories of this property are the time-course of $\mathbf{w}^\top\mathbf{x}$ itself, the instantaneous amplitude (or bandpower) of $\mathbf{w}^\top\mathbf{x}$, and thirdly the instantaneous phase of $\mathbf{w}^\top\mathbf{x}$. As for types of additional information provided to the algorithm, we have binary labels, a univariate target signal, and thirdly an additional multivariate signal.

The first row of this table 7.2 shows the familiar methods LDA, regression and CCA/PLS, which are all defined with respect to the projected signal directly and not to any particular function of the projected signal. These classical methods optimize spatial filters using increasingly complex information when going from left to right in the table. The next row in the table lists the counterparts of the classical methods that are defined with respect to the phase of the projected signal. Here we have Common Spatial Interaction Patterns (CSIP, Meinecke, 2011) that maximizes phase differences between two conditions (or classes) and a method that maximizes the imaginary part of coherence (maxImC, Ewald et al., 2012) between the projections of two datasets. Note that there is currently no phase-counterpart to regression. However, such method can be constructed as a special case of maxImC, just like regression can be constructed as a special case of CCA. Finally, in the third row we come to the decomposition methods that are designed with respect to amplitude dynamics of oscillatory signals. Here we find the familiar CSP algorithm that was reviewed in chapter 2.3 and the members of the SPoC framework.

		signal aspect of interest (dataset $\mathbf{y}$)		
		$\mathbf{w}_\mathbf{y}^\top \mathbf{y}$	Phase $\left(\mathbf{w}_\mathbf{y}^\top \mathbf{y}\right)$	Ampl $\left(\mathbf{w}_\mathbf{y}^\top \mathbf{y}\right)$
signal aspect of interest (dataset $\mathbf{x}$)	$\mathbf{w}_\mathbf{x}^\top \mathbf{x}$	CCA, PLS	—	—
	Phase $\left(\mathbf{w}_\mathbf{x}^\top \mathbf{x}\right)$	?	MaxImC	—
	Ampl $\left(\mathbf{w}_\mathbf{x}^\top \mathbf{x}\right)$	mSPoC	?	cSPoC, Power-CCA

Table 7.3: A categorization of decomposition methods that simultaneously extract components from two datasets ($\mathbf{x}$ and $\mathbf{y}$) according to which aspect of the projected signals ($\mathbf{w}_\mathbf{x}^\top \mathbf{x}$ and $\mathbf{w}_\mathbf{y}^\top \mathbf{y}$) is of interest for the decomposition.

Note that although mSPoC and cSPoC are designed for different purposes, they ended up in the same cell of table 7.2. In order to give a more fine-grained picture of table 7.2's right-most column we have organized the methods of this column into a separate categorization shown in table 7.3. Here we categorize methods that fuse two datasets according to the signal aspect of interest from either of the datasets. Table 7.3 clearly differentiates mSPoC from cSPoC and from the classical methods CCA and PLS. Additionally, this table reveals an interesting point, namely the non-existence of a method that optimizes phase-amplitude-coupling, although this particular phenomenon is the subject of an ongoing research endeavor in the neurosciences (Canolty et al., 2006; Nikulin et al., 2007; Canolty and Knight, 2010; Florin and Baillet, 2015).

Conclusion

With this thesis we have contributed a set of decomposition (or factor-) methods, specifically designed for the extraction of amplitude modulated oscillatory components from high-dimensional uni- and multimodal neuroimaging recordings. We have studied the strength and weaknesses of the proposed methods using theoretical considerations as well as numerical simulations. Additionally, we have demonstrated the practical utility of our methods in a multitude of different real-world neuroimaging datasets recorded with a total of four measurement modalities (EEG, MEG, fNIRS, and fMRI).

The results obtained on the real-world datasets have been in line with what was expected beforehand or with findings confirmed by the literature. One could argue, that therefore our findings have little to none implications for neuroscience as such. While at first glance this may be the case, we firmly believe our results have strong implications with respect to the validity of our methodological developments. In fact, we believe we have presented compelling evidence for the correctness of our approach. This in turn makes our methods attractive choices for future studies of amplitude modulated neural sources, their relation to other aspects of brain activity in within-subject analysis or to corresponding activity from other brains in across-subject analysis, studied in academical, commercial, or clinical settings.

Appendix A

Tricks of the Trade

A.1 Associativity of linear operations

A basic, yet important, property of linear operations is *associativity*. The associative property, applied to matrix multiplication, states that

$$\mathbf{A}\,(\mathbf{BC}) = (\mathbf{AB})\,\mathbf{C}$$

for appropriately sized matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$. In the present context, this property is used to simplify an optimization procedure and at the same time retain applicability of the resulting spatial filters to the original input data.

The optimization of spatial filters can be simplified by first mapping the input data into a space that is more beneficial for optimization compared to the original input space. Such a space could, for example, be of lower dimensionality than the original input space. After optimization in the simplified data space, the results have to be transformed back into the original input space.

The following linear transformation maps the input data into a space spanned by the columns of the matrix $\mathbf{M}$:

$$\widetilde{\mathbf{X}} \quad = \quad \mathbf{M}^\top \mathbf{X}\,. \tag{A.1}$$

The transformed dataset $\widetilde{\mathbf{X}}$ is then used to optimize a given objective function which in turn results in a set of spatial filters $\widetilde{\mathbf{W}}$. In order to apply the obtained spatial filters to the original input dataset, they first have to be projected into the basis that spans the original input dataset. The corresponding projection is achieved by multiplying $\mathbf{M}$ with $\widetilde{\mathbf{W}}$. In order to demonstrate this, let $\mathbf{S}$ denote the signal of interest, which is extracted by means of $\widetilde{\mathbf{W}}$ from $\widetilde{\mathbf{X}}$. Then we have

$$\mathbf{S} = \widetilde{\mathbf{W}}^\top \widetilde{\mathbf{X}} = \widetilde{\mathbf{W}}^\top \left(\mathbf{M}^\top \mathbf{X}\right) = \left(\widetilde{\mathbf{W}}^\top \mathbf{M}^\top\right) \mathbf{X} = \mathbf{W}^\top \mathbf{X} \tag{A.2}$$

where we have simply substituted $\mathbf{M}^\top\mathbf{X}$ for $\widetilde{\mathbf{X}}$, applied the associativity rule, and defined $\mathbf{W} := \mathbf{M}\widetilde{\mathbf{W}}$.

This result may seem rather trivial at first sight. However, the relation above forms the conceptual basis of powerful optimization tricks such as deflation and dimensionality reduction which are introduced below, and thus it is worth including it in this list of tricks.

A.2 Whitening

A data space often beneficial for optimization is one in which all covariances between input variables have been removed. The transformation that maps the original input data into such a space is called *whitening* and the previously introduced PCA can be used to construct a whitening transformation.

Let $\mathbf{V}$ denote the matrix that contains the eigenvectors of the data covariance matrix in its columns, i.e. it holds that

$$\mathbf{C} = \mathbf{V}\Lambda\mathbf{V}^\top ,$$

where the matrix Λ contains the corresponding eigenvectors on the diagonal and only zeros on the off-diagonals. $\mathbf{V}$ is orthonormal, which means that $\mathbf{V}^\top\mathbf{V} = \mathbf{I}$. We can now define the *whitening matrix*

$$\mathbf{M} := \mathbf{V}\Lambda^{-\frac{1}{2}}$$

which contains the whitening filters in its columns. The whitened data is obtained by applying the whitening transform to the original data matrix

$$\widetilde{\mathbf{X}} = \mathbf{M}^\top\mathbf{X} .$$

In order to demonstrate the result of this operation, let us inspect the covariance matrix of $\widetilde{\mathbf{X}}$, which is given by

$$\begin{aligned}
\mathbf{C}_{\widetilde{x}\widetilde{x}} &= \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top \\
&= \mathbf{M}^\top\mathbf{X}\mathbf{X}^\top\mathbf{M} \\
&= \mathbf{M}^\top\mathbf{C}\mathbf{M} \\
&= \Lambda^{-\frac{1}{2}}\mathbf{V}^\top\mathbf{V}\Lambda\mathbf{V}^\top\mathbf{V}\Lambda^{-\frac{1}{2}} \\
&= \mathbf{I} .
\end{aligned} \tag{A.3}$$

The derivation shows that the rows of $\widetilde{\mathbf{X}}$ are uncorrelated.

A.3 Deflation

In some optimization scenarios a set of $K > 1$ components can only be obtained iteratively, i.e. one after the other. In particular, this is the case if the objective function, here denoted by L, can be defined in terms of a single spatial filter only, i.e. if it is of the form

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}}\ L(\mathbf{w}, \mathbf{X}) . \tag{A.4}$$

Let us assume the first component has been found by optimizing the objective function which yielded the spatial filter $\mathbf{w}^1$. Re-running the optimization with the same data will simply yield the same weight vector again. Thus, in order to obtain a component that is different from the first, a constraint has to be introduced that avoids previously obtained solutions. One constraint that is often applied in this context is *mutual (temporal) decorrelation*,

which means that time courses of the components extracted from a single dataset are all decorrelated from each other. In fact, all of the methods discussed in thesis make this assumption.

Let the row vectors $\mathbf{s}^i \in \mathbb{R}^T$ and $\mathbf{s}^j \in \mathbb{R}^T$ denote two different components extracted from $\mathbf{X}$, with $\mathbf{s}^i = \mathbf{w}^{i\top}\mathbf{X}$ and $\mathbf{s}^j = \mathbf{w}^{j\top}\mathbf{X}$. Assuming zero mean, the covariance between them is given by

$$\mathrm{Cov}(\mathbf{s}^i, \mathbf{s}^j) \propto \mathbf{s}^i \mathbf{s}^{j\top} = \mathbf{w}^{i\top}\mathbf{X}\mathbf{X}^\top \mathbf{w}^j = \mathbf{w}^{i\top}\mathbf{C}\mathbf{w}^j \ . \tag{A.5}$$

Since zero covariance implies zero correlation, the decorrelation constraint can be expressed in terms of zero covariance. Thus, when optimizing the i-th spatial filter, with $1 < i \leq K$, the mutual decorrelation constraint is given by demanding $\mathbf{w}^{i\top}\mathbf{C}\mathbf{w}^j = 0$, for all $j < i$.

A particularly convenient way of enforcing this constraint is described next, using the framework of *matrix deflation* (e.g. Mackey, 2009).

First of all, the data has to be whitened. As outlined above, the data is projected using a whitening $\mathbf{M}$ to yield $\widetilde{\mathbf{X}} = \mathbf{M}^\top\mathbf{X}$ and weight vectors are now optimized using $\widetilde{\mathbf{X}}$. This transforms the decorrelation constraint into an *orthogonality constraint*. Thus we now demand $\widetilde{\mathbf{w}}^{i\top}\widetilde{\mathbf{w}}^j = 0$, for all $j < i$, because after whitening the covariance matrix is simply the identity matrix.

Then, after having obtained $\widetilde{\mathbf{w}}^1$ from

$$\widetilde{\mathbf{w}}^1 = \underset{\widetilde{\mathbf{w}}}{\mathrm{argmin}}\ L(\widetilde{\mathbf{w}}, \widetilde{\mathbf{X}}) \ , \tag{A.6}$$

a set of vectors has to be found that together with $\widetilde{\mathbf{w}}^1$ constitute an orthonormal basis of $\mathbb{R}^{N_x}$. Note that such a set is easily obtained using the well-known Gram-Schmidt orthonormalization. Now let the matrix $\mathbf{B}_{\perp 1} \in \mathbb{R}^{N_x \times N_x - 1}$ denote the matrix that contains this set of vectors in its columns. By design we have ${\mathbf{B}_{\perp 1}}^\top \mathbf{B}_{\perp 1} = \mathbf{I}$ and ${\mathbf{B}_{\perp 1}}^\top \widetilde{\mathbf{w}}^1 = \mathbf{0}$. We now project the whitened data into the $N_x - 1$ dimensional space spanned by the columns of $\mathbf{B}_{\perp 1}$ and obtain

$$\widetilde{\widetilde{\mathbf{X}}} = {\mathbf{B}_{\perp 1}}^\top \widetilde{\mathbf{X}} \ , \tag{A.7}$$

which we use to optimize the next spatial filter by means of

$$\widetilde{\widetilde{\mathbf{w}}}^2 = \underset{\widetilde{\widetilde{\mathbf{w}}}}{\mathrm{argmin}}\ L(\widetilde{\widetilde{\mathbf{w}}}, \widetilde{\widetilde{\mathbf{X}}}) \ . \tag{A.8}$$

Note that $\widetilde{\widetilde{\mathbf{w}}}^2$ was optimized in a $N_x - 1$ dimensional space. It is projected to the whitened $\mathbb{R}^{N_x}$ via $\widetilde{\mathbf{w}}^2 := \mathbf{B}_{\perp 1}\widetilde{\widetilde{\mathbf{w}}}^2$

It is now easily verified that $\widetilde{\mathbf{w}}^1$ and $\widetilde{\mathbf{w}}^2$ are indeed orthogonal, because it holds that

$$\widetilde{\mathbf{w}}^{1\top}\widetilde{\mathbf{w}}^2 = \widetilde{\mathbf{w}}^{1\top}\mathbf{B}_{\perp 1}\widetilde{\widetilde{\mathbf{w}}}^2 = \mathbf{0}^\top\widetilde{\widetilde{\mathbf{w}}}^2 = 0. \tag{A.9}$$

Firstly, the decorrelation constrained is transformed into an orthogonality constrained by means of whitening. Thereafter, the orthogonality of subsequent spatial filters is guarantied

by first projecting the data into a subspace that is orthogonal to the so far obtained filters and then optimizing for the next filter within this subspace.

Finally, we show that orthogonality of filters in whitened space implies decorrelation of the projected signals in the un-whitened input space. To see this, we start from $\mathrm{Cov}(\mathbf{s}^i, \mathbf{s}^j) \propto \mathbf{w}^{i\top}\mathbf{C}\mathbf{w}^j$, which was established in Eq. (A.5), and we use the relations $\mathbf{w} = \mathbf{M}\widetilde{\mathbf{w}}$, $\mathbf{M} = \mathbf{V}\Lambda^{-\frac{1}{2}}$, and $\mathbf{C} = \mathbf{V}\Lambda\mathbf{V}^\top$, which were introduced above. Then we have

$$\begin{aligned}
\mathrm{Cov}(\mathbf{s}^i, \mathbf{s}^j) &\propto \mathbf{w}^{i\top}\mathbf{C}\mathbf{w}^j \\
&= \widetilde{\mathbf{w}}^{i\top}\mathbf{M}^\top\mathbf{C}\mathbf{M}\widetilde{\mathbf{w}}^j \\
&= \widetilde{\mathbf{w}}^{i\top}\Lambda^{-\frac{1}{2}\top}\mathbf{V}^\top\mathbf{V}\Lambda\mathbf{V}^\top\mathbf{V}\Lambda^{-\frac{1}{2}}\widetilde{\mathbf{w}}^j \\
&= \widetilde{\mathbf{w}}^{i\top}\Lambda^{-\frac{1}{2}\top}\Lambda\Lambda^{-\frac{1}{2}}\widetilde{\mathbf{w}}^j \\
&= \widetilde{\mathbf{w}}^{i\top}\widetilde{\mathbf{w}}^j \ . \qquad \text{(A.10)}
\end{aligned}$$

This derivation shows that if the filters in whitened space are orthogonal, i.e. if $\widetilde{\mathbf{w}}^{i\top}\widetilde{\mathbf{w}}^j = 0$, then the corresponding component time-courses will be decorrelated. The deflation scheme ensures $\widetilde{\mathbf{w}}^{i\top}\widetilde{\mathbf{w}}^j = 0$ by design, and thereby also uncorrelatedness of extracted components.

Appendix B

hyperSPoC

In the previous chapters we have outlined a number of ways to implement the idea of source power co-modulation, for example via optimization of covariance or correlation. Starting from the basic SPoC idea, we have shown how to extend the basic SPoC algorithms in several multivariate and multimodal ways, leading up to the mSPoC and cSPoC algorithms. In this section, we present a unifying framework that contains all the previously shown SPoC versions as special cases, as well as other well known algorithms. We refer to this approach as *hyperSPoC*.

The hyperSPoC objective function models the co-modulation between three categories of datasets: (i) oscillatory datasets for which the amplitude dynamics are the aspect of interest, (ii) non-oscillatory datasets for which the time-courses of components are the aspect of interest, and (iii) a uni-variate target signal. Also, here we impose no restrictions on the number of datasets within the first two categories of datasets.

Let us now define variables that represent the groups of datasets. Let $\mathbb{X} = \left\{\mathbf{x}^1, ..., \mathbf{x}^{N_\mathbb{X}}\right\}$ denote the set that contains $N_\mathbb{X} \geq 0$ datasets from the first category, $\mathbb{Y} = \left\{\mathbf{y}^1, ..., \mathbf{y}^{N_\mathbb{Y}}\right\}$ denote the set that contains $N_\mathbb{Y} \geq 0$ datasets from the second category, and finally let $\mathbb{Z}$ denote the set that is either empty or contains the target variable z, i.e. we have either $N_\mathbb{Z} = 0$ or $N_\mathbb{Z} = 1$. Because we are going to look at all pairwise combinations of the datasets contained in the three categories, it is convenient to introduce a single ordered set that contains all of the datasets. Let $\mathbb{D} = \left\{\mathbf{x}^1, ..., \mathbf{x}^{N_\mathbb{X}}, \mathbf{y}^1, ..., \mathbf{y}^{N_\mathbb{Y}}, z^{N_\mathbb{Z}}\right\}$ denote this set of datasets, where the cardinality of $\mathbb{D}$ is given by $|\mathbb{D}| = N_\mathbb{D} = N_\mathbb{X} + N_\mathbb{Y} + N_\mathbb{Z} \geq 0$. Finally let $\mathbf{d}^i \in \mathbb{D}$, for $i \in \{1, ..., N_\mathbb{D}\}$ denote a single dataset from the ordered set $\mathbb{D}$ where its category depends on the index i.

In analogy to the other methods presented before, we are going to extract components from the datasets by using spatial filters. Thus, the objective function will be parametrized by a set of weight vectors $\mathbf{w}^i$ for $i \in \{1, ..., N_\mathbb{D}\}$ and the projected signals are denoted by ${\mathbf{w}^i}^\top \mathbf{d}^i$. An exception has to be made for the uni-variate target function z which, if present, is also an element of $\mathbb{D}$. In this case the respective 'weight vector' has only one entry which we fix to the constant 1.

The next step consists of defining fixed functions that extract the relevant aspect of interest from the projected signals. Let us denote this feature extraction function by $\Phi_i\left({\mathbf{w}^i}^\top \mathbf{d}^i\right)$, where the index i determines the semantics of Φ_i depending on what category the dataset $\mathbf{d}^i$ belongs to. For datasets of the first category, the relevant signal aspect is amplitude

dynamics and thus Φ_i is defined to be the envelope of the projected signal, computed using the Hilbert transform. See section 6.1.1 for more details on the definition of the envelope. For signals of category two and three, Φ_i is simply the identity function.

The last elements that we need for the objective function are the co-modulation function and weighting parameters that can be set to give more or less preference to certain dataset combinations. Here we choose correlation as a measure of co-modulation and we define the scalar constants $c_{i,j} \in \mathbb{R}$ for $1 \leq i \leq N_{\mathbb{D}}$ and $1 \leq j \leq N_{\mathbb{D}}$. The coefficients $c_{i,j}$ will not be optimized here and in the simplest case can be set to 1 for all i and j.

Equipped with these definitions we can now define an objective function that combines an arbitrary number of datasets from three different categories of datasets by maximizing the pairwise coupling between components. The hyperSPoC objective function is given by

$$\begin{aligned} f &= \sum_{i,j} c_{i,j} \cdot \mathrm{Corr}(\Phi_i, \Phi_j) \\ &= \sum_{i,j} c_{i,j} \cdot \frac{\mathrm{Cov}(\Phi_i, \Phi_j)}{\sqrt{\mathrm{Var}(\Phi_i)\,\mathrm{Var}(\Phi_j)}} \\ &= \sum_{i,j} c_{i,j} \cdot \frac{\langle \Phi_i, \Phi_j \rangle}{\sqrt{\langle \Phi_i^2 \rangle \langle \Phi_j^2 \rangle}} \,, \end{aligned} \tag{B.1}$$

where we have used some abbreviations for the sake of readability: $\sum_{i,j}$ is short-hand for the double sum $\sum_{i=1}^{N_{\mathbb{D}}} \sum_{j=1, j\neq i}^{N_{\mathbb{D}}}$ and Φ_i is short for $\Phi_i\left(\mathbf{w}^{i\top}\mathbf{d}^i\right)$. Thus, keep in mind that Φ_i is a function of $\mathbf{w}^i$.

Let us take a moment to go through possible combinations of 'dataset scenarios' in order to see what cases this objective function covers. Remember that the variables $N_{\mathbb{X}}$, $N_{\mathbb{Y}}$, and $N_{\mathbb{Z}}$ respectively denote the number of datasets from each of the three categories defined earlier. Table B.1 lists a number of dataset combinations that, when applied to the hyperSPoC objective, lead to algorithms previously discussed in this thesis.

dataset type combinations	resulting algorithm	notes
$N_{\mathbb{X}} = 0, N_{\mathbb{Y}} = 1, N_{\mathbb{Z}} = 1$	LDA or regression	the type of target variable (binary or not) decides
$N_{\mathbb{X}} = 0, N_{\mathbb{Y}} \geq 2, N_{\mathbb{Z}} = 0$	(multiset-)CCA/PLS	pre-whitening leads to PLS
$N_{\mathbb{X}} = 1, N_{\mathbb{Y}} = 0, N_{\mathbb{Z}} = 1$	CSP or SPoC	the type of target variable (binary or not) decides
$N_{\mathbb{X}} = 1, N_{\mathbb{Y}} = 1, N_{\mathbb{Z}} = 0$	mSPoC	temporal shift is not explicitly modeled
$N_{\mathbb{X}} \geq 2, N_{\mathbb{Y}} = 0, N_{\mathbb{Z}} = 0$	(multiset-)cSPoC	

Table B.1: A listing of algorithms that result from the hyperSPoC objective for specific combinations of input datasets.

The hyperSPoC objective has no analytic solution. However, the gradient with respect to all weight vectors $\mathbf{w}^k$ can be computed and then it can be optimized using numerical optimization packages. For the gradient with respect to $\mathbf{w}^k$ we have

$$\frac{\partial f}{\partial \mathbf{w}^k} = \sum_{i,j} c_{i,j} \cdot \frac{\partial \operatorname{Corr}(\Phi_i, \Phi_j)}{\partial \mathbf{w}^k}. \tag{B.2}$$

In order to simplify subsequent derivations, let us define

$$\begin{aligned} \operatorname{Corr}(\Phi_i, \Phi_j) &= \frac{\langle \Phi_i, \Phi_j \rangle}{\sqrt{\langle \Phi_i^2 \rangle \langle \Phi_j^2 \rangle}} \\ &:= \frac{g}{h}. \end{aligned} \tag{B.3}$$

The partial derivatives of g and h with respect to $\mathbf{w}^k$ are as follows:

$$\frac{\partial g}{\partial \mathbf{w}^k} = \left\langle \frac{\partial \Phi_i}{\partial \mathbf{w}^k} \Phi_j \right\rangle + \left\langle \frac{\partial \Phi_j}{\partial \mathbf{w}^k} \Phi_i \right\rangle \tag{B.4}$$

$$\frac{\partial h}{\partial \mathbf{w}^k} = \frac{1}{h} \left(\left\langle \frac{\partial \Phi_i}{\partial \mathbf{w}^k} \Phi_i \right\rangle \langle \Phi_j^2 \rangle + \left\langle \frac{\partial \Phi_j}{\partial \mathbf{w}^k} \Phi_j \right\rangle \langle \Phi_i^2 \rangle \right). \tag{B.5}$$

We can now derive the partial derivative of each summand in Eq. (B.2)

$$\begin{aligned} \frac{\partial \operatorname{Corr}(\Phi_i, \Phi_j)}{\partial \mathbf{w}^k} &= \frac{\partial \frac{g}{h}}{\partial \mathbf{w}^k} \\ &= \frac{\frac{\partial g}{\partial \mathbf{w}^k} h - g \frac{\partial h}{\partial \mathbf{w}^k}}{h^2} \\ &= \frac{1}{h} \frac{\partial g}{\partial \mathbf{w}^k} - \frac{g}{h^2} \frac{\partial h}{\partial \mathbf{w}^k} \\ &= \frac{1}{h} \left(\left\langle \frac{\partial \Phi_i}{\partial \mathbf{w}^k} \Phi_j \right\rangle + \left\langle \frac{\partial \Phi_j}{\partial \mathbf{w}^k} \Phi_i \right\rangle \right) \\ &\quad - \frac{g}{h^3} \left(\left\langle \frac{\partial \Phi_i}{\partial \mathbf{w}^k} \Phi_i \right\rangle \langle \Phi_j^2 \rangle + \left\langle \frac{\partial \Phi_j}{\partial \mathbf{w}^k} \Phi_j \right\rangle \langle \Phi_i^2 \rangle \right) \\ &= \frac{1}{h} \left\langle \Phi_i \underbrace{\left(\frac{\partial \Phi_j}{\partial \mathbf{w}^k} - \frac{\partial \Phi_i}{\partial \mathbf{w}^k} \frac{\langle \Phi_i \Phi_j \rangle}{\langle \Phi_i^2 \rangle} \right)}_{:= \Delta_{kji}} \right\rangle \\ &\quad + \frac{1}{h} \left\langle \Phi_j \underbrace{\left(\frac{\partial \Phi_i}{\partial \mathbf{w}^k} - \frac{\partial \Phi_j}{\partial \mathbf{w}^k} \frac{\langle \Phi_i \Phi_j \rangle}{\langle \Phi_j^2 \rangle} \right)}_{:= \Delta_{kij}} \right\rangle \\ &= \frac{\langle \Phi_i \Delta_{kji} \rangle}{\sqrt{\langle \Phi_i^2 \rangle \langle \Phi_j^2 \rangle}} + \frac{\langle \Phi_j \Delta_{kij} \rangle}{\sqrt{\langle \Phi_i^2 \rangle \langle \Phi_j^2 \rangle}} \end{aligned} \tag{B.6}$$

What is left at this point, is to compute $\frac{\partial \Phi_i}{\partial \mathbf{w}^k}$ for all i. For datasets of type two and three, this is trivial. See section 6.1.1 for the partial derivative for datasets of type three, where Φ_i corresponds to the extraction of the envelope.

Bibliography

Akaike, H. (1976). Canonical correlation analysis of time series and the use of an information criterion. In Mehra, R. K. and Lainiotis, D. G., editors, *System Identification Advances and Case Studies*, volume 126 of *Mathematics in Science and Engineering*, pages 27 – 96. Elsevier.

Arthurs, O. J. and Boniface, S. (2002). How well do we understand the neural origins of the fMRI BOLD signal? *Trends in Neurosciences*, 25(1):27–31.

Astolfi, L., Cincotti, F., Mattia, D., Marciani, M. G., Baccala, L. A., de Vico Fallani, F., Salinari, S., Ursino, M., Zavaglia, M., Ding, L., Edgar, J. C., Miller, G. A., He, B., and Babiloni, F. (2007). Comparison of different cortical connectivity estimators for high-resolution eeg recordings. *Human Brain Mapping*, 28(2):143–157.

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

Başar, E., Schürmann, M., Başar-Eroglu, C., and Karakaş, S. (1997). Alpha oscillations in brain functioning: an integrative theory. *International Journal of Psychophysiology*, 26(1-3):5–29.

Baillet, S., Mosher, J. C., and Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, 18(6):14–30.

Barlow, J. S. (1993). *The Electroencephalogram: Its Patterns and Origins*. The MIT Press.

Bauer, M., Oostenveld, R., Peeters, M., and Fries, P. (2006). Tactile spatial attention enhances gamma-band activity in somatosensory cortex and reduces low-frequency activity in parieto-occipital areas. *The Journal of Neuroscience*, 26(2):490–501.

Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159.

Belouchrani, A., Abed-Meraim, K., Cardoso, J.-F., and Moulines, E. (1997). A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444.

Berger, H. (1929). Über das Elektroenkephalogramm des Menschen. *Archiv fur Psychiatrie und Nervenkrankheiten*, 87:527–570.

Berwick, J., Johnston, D., Jones, M., Martindale, J., Martin, C., Kennerley, A., Redgrave, P., and Mayhew, J. (2008). 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):787–798.

Bie, T. D. and Moor, B. D. (2003). On the regularization of canonical correlation analysis. *International Symposium on Independent Component Analysis and Blind Signal Separation*.

Bießmann, F., Meinecke, F. C., Gretton, A., Rauch, A., Rainer, G., Logothetis, N., and Müller, K.-R. (2009). Temporal kernel canonical correlation analysis and its application in multimodal neuronal data analysis. *Machine Learning*, 79(1-2):5—27.

Bießmann, F., Murayama, Y., Logothetis, N. K., Müller, K.-R., and Meinecke, F. C. (2012). Improved decoding of neural activity from fMRI signals using non-separable spatiotemporal deconvolutions. *NeuroImage*, 61(4):1031–1042.

Bießmann, F., Plis, S. M., Meinecke, F. C., Eichele, T., and Müller, K.-R. (2011). Analysis of multimodal neuroimaging data. *Biomedical Engineering, IEEE Reviews in*, 4:26–58.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., and Curio, G. (2007). The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37(2):539–550.

Blankertz, B., Lemm, S., Treder, M. S., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of ERP components – a tutorial. *NeuroImage*, 56:814–825.

Blankertz, B., Sannelli, C., Halder, S., Hammer, E. M., Kübler, A., Müller, K.-R., Curio, G., and Dickhaus, T. (2010). Neurophysiological predictor of SMR-based BCI performance. *NeuroImage*, 51(4):1303–1309.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Müller, K.-R. (2008). Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, 25(1):41–56.

Bonvento, G., Sibson, N., and Pellerin, L. (2002). Does glutamate image your thoughts? *Trends in Neurosciences*, 25(7):359–364.

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience*, 16(13):4207–4221.

Bressler, S. L. and Seth, A. K. (2011). Wiener-Granger Causality: A well established methodology. *Neuroimage*, 58(2):323–329.

Buxton, R. B., Uludağ, K., Dubowitz, D. J., and Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *NeuroImage*, 23:S220–S233.

Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents-EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13(6):407–420.

Buzsáki, G. and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679):1926–1929.

Calhoun, V. D., Adali, T., Pearlson, G., and Kiehl, K. (2006). Neuronal chronometry of target detection: fusion of hemodynamic and event-related potential data. *NeuroImage*, 30(2):544–553.

Calhoun, V. D., Liu, J., and Adalı, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and erp data. *NeuroImage*, 45(1):S163–S172.

Campi, C., Parkkonen, L., Hari, R., and Hyvärinen, A. (2013). Non-linear canonical correlation for joint analysis of MEG signals from two subjects. *Frontiers in Neuroscience*, 7(107).

Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science*, 313(5793):1626–1628.

Canolty, R. T. and Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends in Cognitive Sciences*, 14(11):506 – 515.

Carbonell, F., Worsley, K. J., Trujillo-Barreto, N. J., and Sotero, R. C. (2009). Random fields—union intersection tests for detecting functional connectivity in EEG/MEG imaging. *Human Brain Mapping*, 30(8):2477–2486.

Cardoso, J.-F. and Souloumiac, A. (1993a). Blind beamforming for non-gaussian signals. *IEE Proceedings F (Radar and Signal Processing)*, 140(6):362–370.

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

Cohen, D. (1968). Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents. *Science*, 161(3843):784–786.

Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, 36(3):287–314.

Correa, N., Li, Y.-O., Adali, T., and Calhoun, V. (2009). Fusion of fMRI, sMRI, and EEG data using canonical correlation analysis. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009*, pages 385–388.

Correa, N. M., Adali, T., Li, Y.-O., and Calhoun, V. D. (2010). Canonical correlation analysis for data fusion and group inferences. *IEEE Signal Processing Magazine*, 27(4):39–50.

Correa, N. M., Li, Y.-O., Adali, T., and Calhoun, V. D. (2008). Canonical correlation analysis for feature-based fusion of biomedical imaging modalities and its application to detection of associative networks in schizophrenia. *IEEE Journal of Selected Topics in Signal Processing*, 2(6):998–1007.

Dähne, S., Bießman, F., Meinecke, F. C., Mehnert, J., Fazli, S., and Müller, K.-R. (2013). Integration of multivariate data streams with bandpower signals. *IEEE Transactions on Multimedia*, 15(5):1001–1013.

Dähne, S., Bießmann, F., Meinecke, F. C., Mehnert, J., Fazli, S., and Müller, K.-R. (2014a). Multimodal integration of electrophysiological and hemodynamic signals. In *Brain-Computer Interface (BCI), 2014 International Winter Workshop on*, pages 1–4.

Dähne, S., Bießmann, F., Samek, W., Haufe, S., Goltz, D., Gundlach, C., Villringer, A., Fazli, S., and Müller, K.-R. (2015). Multivariate machine learning methods for fusing functional multimodal neuroimaging data. *Proceedings of the IEEE*, 103(9):1507 – 1530.

Dähne, S., Hahne, J., Pawletta, P., and Müller, K.-R. (2014b). Boosting simultaneous and proportional myoelectric control by combining source power correlation (spoc) and linear regression. In *Bernstein Conference, 2014*.

Dähne, S., Haufe, S., Bießmann, F., Meinecke, F., Ramirez, D., Schreier, P., Nikulin, V., and Müller, K.-R. (2014c). Finding brain oscillations with power dependencies in neuroimaging data. In *Annual Meeting of the Organization for Human Brain Mapping (OHBM), 2014*.

Dähne, S., Höhne, J., Schreuder, M., and Tangermann, M. (2011a). Band power features correlate with performance in auditory brain-computer interface. In *Front. Hum. Neurosci. Conference Abstract: XI International Conference on Cognitive Neuroscience (ICON XI)*.

Dähne, S., Höhne, J., Schreuder, M., and Tangermann, M. (2011b). Slow feature analysis - a tool for extraction of discriminating event-related potentials in brain-computer interfaces. In Honkela, T., Duch, W., Girolami, M., and Kaski, S., editors, *Artificial Neural Networks and Machine Learning - ICANN 2011*, volume 6791 of *Lecture Notes in Computer Science*, pages 36–43. Springer Berlin / Heidelberg.

Dähne, S., Höhne, J., and Tangermann, M. (2011c). Adaptive classification improves control performance in ERP-based BCIs. In *Proceedings of the 5th International BCI Conference*, pages 92–95, Graz.

Dähne, S., Meinecke, F., Haufe, S., Höhne, J., Tangermann, M., Nikulin, V., and Müller, K.-R. (2012). Multi-variate correlation of power spectral density. In *Annual Meeting of the Organization for Human Brain Mapping (OHBM), 2012*.

Dähne, S., Meinecke, F. C., Haufe, S., Höhne, J., Tangermann, M., Müller, K.-R., and Nikulin, V. V. (2014d). SPoC: a novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. *NeuroImage*, 86(0):111–122.

Dähne, S., Nikulin, V. V., Ramírez, D., Schreier, P. J., Müller, K.-R., and Haufe, S. (2014e). Finding brain oscillations with power dependencies in neuroimaging data. *NeuroImage*, 96:334–348.

Dähne, S., Nikulin, V. V., Ramirez, D., Schreier, P. J., Müller, K.-R., and Haufe, S. (2014f). Optimizing spatial filters for the extraction of envelope-coupled neural oscillations. In *International Workshop on Pattern Recognition in Neuroimaging, 2014*. IEEE.

Dähne, S., Wilbert, N., and Wiskott, L. (2014g). Slow feature analysis of retinal waves leads to V1 complex cells. *PLoS Computational Biology*, 10(5):e1003564.

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

Darchia, N., Campbell, I., Tan, X., and Feinberg, I. (2007). Kinetics of NREM delta EEG power density across NREM periods depend on age and on delta-band designation. *Sleep*, 30(1):71–79.

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

Debener, S., Herrmann, C. A. C. S., Kranczioch, C., Gembris, D., and Engel, A. K. (2003). Top-down attentional processing enhances auditory evoked gamma band activity. *NeuroReport*, 14(5):683–686.

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

Delpy, D. T., Cope, M., van der Zee, P., Arridge, S., Wray, S., and Wyatt, J. (1988). Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in Medicine and Biology*, 33(12):1433–1442.

Demanuele, C., Broyd, S. J., Sonuga-Barke, E. J. S., and James, C. (2012). Neuronal oscillations in the EEG under varying cognitive load: A comparative study between slow waves and faster oscillations. *Clinical Neurophysiology*, 124(2):247–262.

Denker, M., Roux, S., Lindén, H., Diesmann, M., Riehle, A., and Grün, S. (2011). The local field potential reflects surplus spike synchrony. *Cerebral Cortex*, 21(12):2681–2695.

Devor, A., Ulbert, I., Dunn, A. K., Narayanan, S. N., Jones, S. R., Andermann, M. L., Boas, D. A., and Dale, A. M. (2005). Coupling of the cortical hemodynamic response to cortical and thalamic neuronal activity. *Proceedings of the National Academy of Sciences*, 102(10):3822–3827.

Dyrholm, M., Makeig, S., and Hansen, L. K. (2007). Model selection for convolutive ICA with an application to spatiotemporal analysis of EEG. *Neural Computation*, 19:934–955.

Eichele, T., Specht, K., Moosmann, M., Jongsma, M. L. A., Quiroga, R. Q., Nordby, H., and Hugdahl, K. (2005). Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional mri. *Proceedings of the National Academy of Sciences*, 102(49):17798–803.

Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M.-H., Evans, A. C., Zilles, K., and Amunts, K. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *NeuroImage*, 36(3):511–521.

Engel, A. K., Gerloff, C., Hilgetag, C. C., and Nolte, G. (2013). Intrinsic coupling modes: Multiscale interactions in ongoing brain activity. *Neuron*, 80(4):867–886.

Ewald, A., Marzetti, L., Zappasodi, F., Meinecke, F. C., and Nolte, G. (2012). Estimating true brain connectivity from EEG/MEG data invariant to linear and static transformations in sensor space. *NeuroImage*, 60(1):476 – 488.

Fazli, S., Dähne, S., Samek, W., Bießmann, F., and Müller, K.-R. (2015). Learning from more than one data source: data fusion techniques for sensorimotor rhythm-based Brain-Computer Interfaces. *Proceedings of the IEEE*, 103(6):891–906.

Fazli, S., Mehnert, J., Steinbrink, J., Curio, G., Villringer, A., Müller, K.-R., and Blankertz, B. (2012). Enhanced performance by a Hybrid NIRS-EEG Brain Computer Interface. *NeuroImage*, 59(1):519–529. Open Access.

FitzGerald, T. H., Valentin, A., Selway, R., and Richardson, M. P. (2013). Cross-frequency coupling within and between the human thalamus and neocortex. *Frontiers in Human Neuroscience*, 7.

Florin, E. and Baillet, S. (2015). The brain's resting-state activity is shaped by synchronized cross-frequency coupling of neural oscillations. *NeuroImage*, 111:26–35.

Fonov, V., Evans, A., Botteron, K., Almli, C., McKinstry, R., Collins, D., and BDCG (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54.

Francis, J. G. (1961). The QR transformation a unitary analogue to the lr transformation—part 1. *The Computer Journal*, 4(3):265–271.

Freyer, F., Aquino, K., Robinson, P. A., Ritter, P., and Breakspear, M. (2009). Bistability and non-Gaussian fluctuations in spontaneous cortical activity. *The Journal of Neuroscience*, 29(26):8512–8524.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2(4):189–210.

Furl, N., Coppola, R., Averbeck, B. B., and Weinberger, D. R. (2014). Cross-frequency power coupling between hierarchically organized face-selective areas. *Cerebral Cortex*, 24(9):2409–2420.

Gaebler, M., Bießmann, F., Lamke, J.-P., Müller, K.-R., Walter, H., and Hetzer, S. (2014). Stereoscopic depth increases intersubject correlations of brain networks. *NeuroImage*, 100:427–434.

Galambos, R., Makeig, S., and Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proceedings of the National Academy of Sciences of the United States of America*, 78(4):2643–2647.

Gevins, A., Leong, H., Du, R., Smith, M. E., Le, J., DuRousseau, D., Zhang, J., and Libove, J. (1995). Towards measurement of brain function in operational environments. *Biological Psychology*, 40(1-2):169–186.

Goense, J. and Logothetis, N. K. (2008). Neurophysiology of the BOLD fMRI signal in awake monkeys. *Current Biology*, 18(9):631–640.

Gómez-Herrero, G., Atienza, M., Egiazarian, K., and Cantero, J. L. (2008). Measuring directional coupling between EEG sources. *NeuroImage*, 43:497–508.

Gramfort, A., Strohmeier, D., Haueisen, J., Hämäläinen, M., and Kowalski, M. (2013). Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations. *NeuroImage*, 70(0):410–422.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438.

Grinvald, A., Lieke, E., Frostig, R. D., Gilbert, C. D., and Wiesel, T. N. (1986). Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature*, 324:361–364.

Grosse-Wentrup, M., Scholkopf, B., and Hill, J. (2011). Causal influence of gamma oscillations on the sensorimotor rhythm. *NeuroImage*, 56(2):837–842.

Groves, A. R., Beckmann, C. F., Smith, S. M., and Woolrich, M. W. (2011). Linked independent component analysis for multimodal data fusion. *NeuroImage*, 54(3):2198–2217.

Gutmann, M. U. and Hyvärinen, A. (2011). Extracting coactivated features from multiple data sets. In *Artificial Neural Networks and Machine Learning–ICANN 2011*, pages 323–330. Springer.

Haegens, S., Händel, B. F., and Jensen, O. (2011a). Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. *The Journal of Neuroscience*, 31(14):5197–5204.

Haegens, S., Nácher, V., Hernández, A., Luna, R., Jensen, O., and Romo, R. (2011b). Beta oscillations in the monkey sensorimotor network reflect somatosensory decision making. *Proceedings of the National Academy of Sciences of the United States of America*, 108(26):10708–13.

Hahne, J., Dähne, S., Hwang, H.-J., , Müller, K.-R., and Parra, L. C. (2015). Concurrent adaptation of human and machine improves simultaneous and proportional myoelectric control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. accepted for publication.

Hall, C. N., Reynell, C., Gesslein, B., Hamilton, N. B., Mishra, A., Sutherland, B. A., O'Farrell, F. M., Buchan, A. M., Lauritzen, M., and Attwell, D. (2014). Capillary pericytes regulate cerebral blood flow in health and disease. *Nature*, 508(7494):55–60.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics*, 65(2):413.

Hamel, E. (2006). Perivascular nerves and the regulation of cerebrovascular tone. *Journal of Applied Physiology*, 100(3):1059–1064.

Handwerker, D. A., Ollinger, J. M., and D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, 21(4):1639–1651.

Hansen, S. T., Winkler, I., Hansen, L. K., Müller, K.-R., and Dähne, S. (2015). Fusing simultaneous EEG and fMRI using functional and anatomical information. In *International Workshop on Pattern Recognition in Neuroimaging, 2015*. IEEE. accepted.

Hari, R., Hämäläinen, M., and Joutsiniemi, S. L. (1989). Neuromagnetic steady-state responses to auditory stimuli. *The Journal of the Acoustical Society of America*, 86(3):1033–9.

Harmeling, S., Ziehe, A., Kawanabe, M., and Müller, K.-R. (2003). Kernel-based nonlinear blind source separation. *Neural Computation*, 15:1089–1124.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, 303(5664):1634–1640.

Haufe, S., Dähne, S., and Nikulin, V. V. (2014a). Dimensionality reduction for the analysis of brain oscillations. *NeuroImage*, 101:583–597.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., and Bießmann, F. (2014b). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, 87:96–110. Neuroimage single best paper of 2014 Award.

Haufe, S., Nikulin, V. V., Müller, K.-R., and Nolte, G. (2012). A critical assessment of connectivity measures for EEG data: a simulation study. *NeuroImage*, 64:120–133.

Haufe, S., Nikulin, V. V., Ziehe, A., Müller, K.-R., and Nolte, G. (2008). Combining sparsity and rotational invariance in EEG/MEG source reconstruction. *NeuroImage*, 42(2):726–738.

Haufe, S., Tomioka, R., Dickhaus, T., Sannelli, C., Blankertz, B., Nolte, G., and Müller, K.-R. (2011). Large-scale EEG/MEG source localization with spatial flexibility. *NeuroImage*, 54:851–859.

Haufe, S., Tomioka, R., Nolte, G., Müller, K.-R., and Kawanabe, M. (2010). Modeling sparse connectivity between underlying brain sources for EEG/MEG. *IEEE Transactions on Biomedical Engineering*, 57(8):1954 – 1963.

Heeger, D. J. and Ress, D. (2002). What does fMRI tell us about neuronal activity? *Nature Reviews Neuroscience*, 3(2):142–151.

Herdman, A. T., Lins, O., Van Roon, P., Stapells, D. R., Scherg, M., and Picton, T. W. (2002). Intracerebral sources of human auditory steady-state responses. *Brain Topography*, 15(2):69–86.

Hindriks, R., van Putten, M. J., and Deco, G. (2014). Intra-cortical propagation of EEG alpha oscillations. *NeuroImage*, 103:444–453.

Hipp, J. F., Hawellek, D. J., Corbetta, M., Siegel, M., and Engel, A. K. (2012). Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature Neuroscience*, 15(6):884–890.

Hodgkin, A. L. and Huxley, A. F. (1939). Action potentials recorded from inside a nerve fibre. *Nature*, 144(3651):710–711.

Höhne, J., Krenzlin, K., Dähne, S., and Tangermann, M. (2012). Natural stimuli improve auditory BCIs with respect to ergonomics and performance. *Journal of Neural Engineering*, 9(4):045003.

Holm, A., Lukander, K., Korpela, J., Sallinen, M., and Müller, K. M. I. (2009). Estimating brain load from the EEG. *The Scientific World Journal*, 9.

Horn, F. and Dähne, S. (2013). Increasing the spectral signal-to-noise ratio of common spatial patterns. In *5th International BCI Meeting in Asilomar*.

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

Huster, R. J., Debener, S., Eichele, T., and Herrmann, C. S. (2012). Methods for simultaneous EEG-fMRI: an introductory review. *The Journal of Neuroscience*, 32(18):6053–6060.

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

Hyvärinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492.

Ives, J., Warach, S., Schmitt, F., Edelman, R., and Schomer, D. (1993). Monitoring the patient's EEG during echo planar MRI. *Electroencephalography and Clinical Neurophysiology*, 87(6):417–420.

Jenatton, R., Gramfort, A., Michel, V., Obozinski, G., Eger, E., Bach, F., and Thirion, B. (2012). Multiscale mining of fMRI data with hierarchical structured sparsity. *SIAM Journal on Imaging Sciences*, 5(3):835–856.

Jensen, O. and Colgin, L. L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends in cognitive sciences*, 11(7):267–269.

Jensen, O., Kaiser, J., and Lachaux, J.-P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, 30(7):317–324.

Jobsis, F. F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science*, 198(4323):1264–1267.

John, M. S., Dimitrijevic, A., and Picton, T. W. (2003). Efficient stimuli for evoking auditory steady-state responses. *Ear and Hearing*, 24(5):406–423.

Jolliffe, I. T. (1982). A note on the use of principal components in regression. *Applied Statistics*, pages 300–303.

Jurkiewicz, M. T., Gaetz, W. C., Bostan, A. C., and Cheyne, D. (2006). Post-movement beta rebound is generated in motor cortex: Evidence from neuromagnetic recordings. *NeuroImage*, 32(3):1281–1289.

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

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29(2-3):169–195.

Klostermann, F., Nikulin, V. V., Kühn, A. A., Marzinzik, F., Wahl, M., Pogosyan, A., Kupsch, A., Schneider, G.-H., Brown, P., and Curio, G. (2007). Task-related differential dynamics of EEG alpha-and beta-band synchronization in cortico-basal motor structures. *European Journal of Neuroscience*, 25(5):1604–1615.

Kocsis, L., Herman, P., and Eke, A. (2006). The modified Beer-Lambert law revisited. *Phys Med Biol*, 51:N91–98.

Koles, Z. J. (1991). The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. *Electroencephalography and clinical neurophysiology*, 79(6):440–447.

Krishnan, A., Williams, L. J., McIntosh, A. R., and Abdi, H. (2011). Partial least squares (PLS) methods for neuroimaging: A tutorial and review. *NeuroImage*, 56(2):455–475.

Kublanovskaya, V. N. (1962). On some algorithms for the solution of the complete eigenvalue problem. *USSR Computational Mathematics and Mathematical Physics*, 1(3):637–657.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., and Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, 89(12):5675–5679.

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

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

Lee, H., Kim, Y.-D., Cichocki, A., and Choi, S. (2007). Nonnegative tensor factorization for continuous EEG classification. *International journal of neural systems*, 17(04):305–317.

Lee Rodgers, J. and Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66.

Lemieux, L., Allen, P. J., Franconi, F., Symms, M. R., and Fish, D. K. (1997). Recording of EEG during fMRI experiments: patient safety. *Magnetic Resonance in Medicine*, 38(6):943–952.

Lemm, S., Blankertz, B., Dickhaus, T., and Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *NeuroImage*, 56:387–399.

Litvak, V., Jha, A., Eusebio, A., Oostenveld, R., Foltynie, T., Limousin, P., Zrinzo, L., Hariz, M. I., Friston, K., and Brown, P. (2011). Resting oscillatory cortico-subthalamic connectivity in patients with Parkinson's disease. *Brain*, 134(2):359–374.

Liu, J. and Calhoun, V. (2007). Parallel independent component analysis for multimodal analysis: application to fMRI and EEG data. In *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro 2007*, pages 1028–1031. IEEE.

Liu, J., Pearlson, G., Windemuth, A., Ruano, G., Perrone-Bizzozero, N. I., and Calhoun, V. (2009). Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA. *Human Brain Mapping*, 30(1):241–255.

Liu, Z., de Zwart, J. A., van Gelderen, P., Kuo, L.-W., and Duyn, J. H. (2012). Statistical feature extraction for artifact removal from concurrent fMRI-EEG recordings. *NeuroImage*, 59(3):2073–2087.

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

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

Logothetis, N. K. and Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology*, 66:735–769.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Mackey, L. W. (2009). Deflation methods for sparse PCA. In *Advances in neural information processing systems*, pages 1017–1024.

Maeder, C. L., Sannelli, C., Haufe, S., and Blankertz, B. (2012). Prestimulus sensorimotor rhythms influence brain-computer interface classification performance. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20:653–662.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11:19–60.

Makeig, S. and Jung, T. P. (1996). Tonic, phasic, and transient EEG correlates of auditory awareness in drowsiness. *Cognitive Brain Research*, 4(1):15–25.

Mantini, D., Perrucci, M. G., Del Gratta, C., Romani, G. L., and Corbetta, M. (2007). Electrophysiological signatures of resting state networks in the human brain. *Proceedings of the National Academy of Sciences*, 104(32):13170–13175.

Martel, A., Dähne, S., and Blankertz, B. (2014). EEG predictors of covert vigilant attention. *Journal of Neural Engineering*, 11(3):035009.

Martindale, J., Mayhew, J., Berwick, J., Jones, M., Martin, C., Johnston, D., Redgrave, P., and Zheng, Y. (2003). The hemodynamic impulse response to a single neural event. *Journal of Cerebral Blood Flow & Metabolism*, 23(5):546–555.

Martínez-Montes, E., Valdés-Sosa, P. A., Miwakeichi, F., Goldman, R. I., and Cohen, M. S. (2004). Concurrent EEG/fMRI analysis by multiway partial least squares. *NeuroImage*, 22(3):1023–1034.

McKeown, M. J., Makeig, S., Brown, G. G., Jung, T.-P., Kindermann, S. S., Bell, A. J., and Sejnowski, T. J. (1997). Analysis of fMRI data by blind separation into independent spatial components. Technical report, DTIC Document.

Meinecke, F. (2011). *Identifying Interactions from Superimposed Signals*. PhD thesis, Technische Universtität Berlin.

Meinecke, F. C., Ziehe, A., Kawanabe, M., and Müller, K.-R. (2002). A Resampling Approach to Estimate the Stability of one- or multidimensional Independent Components. *IEEE Transactions on Biomedical Engineering*, 49(12):1514–1525.

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., Wiest, M. C., Karpov, I., King, R. D., Apple, N., et al. (2002). Hyperscanning: simultaneous fMRI during linked social interactions. *NeuroImage*, 16(4):1159–1164.

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

Mørup, M., Hansen, L. K., Arnfred, S. M., Lim, L.-H., and Madsen, K. H. (2008). Shift-invariant multilinear decomposition of neuroimaging data. *NeuroImage*, 42(4):1439–1450.

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

Müller, K.-R., Smola, A., Rätsch, G., Schölkopf, B., Kohlmorgen, J., and Vapnik, V. (1997). Predicting time series with support vector machines. In Gerstner, W., Germond, A., Hasler, M., and Nicoud, J.-D., editors, *Artificial Neural Networks – ICANN '97*, volume 1327 of *LNCS*, pages 999–1004, Berlin. Springer.

Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., and Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. *Journal of neuroscience methods*, 167(1):82–90.

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

Niedermeyer, E. and da Silva, F. L. (2005). *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.

Nierhaus, T., Gundlach, C., Goltz, D., Thiel, S. D., Pleger, B., and Villringer, A. (2013). Internal ventilation system of mr scanners induces specific EEG artifact during simultaneous EEG-fMRI. *NeuroImage*, 74:70–76.

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

Nikulin, V. V., Linkenkaer-Hansen, K., Nolte, G., Lemm, S., Müller, K.-R., Ilmoniemi, R. J., and Curio, G. (2007). A novel mechanism for evoked responses in human brain. *The European journal of neuroscience*, 25:3146–54.

Nikulin, V. V., Nolte, G., and Curio, G. (2011). A novel method for reliable and fast extraction of neuronal EEG/MEG oscillations on the basis of spatio-spectral decomposition. *NeuroImage*, 55(4):1528–1535.

Nikulin, V. V., Nolte, G., and Curio, G. (2012). Cross-frequency decomposition: A novel technique for studying interactions between neuronal oscillations with different frequencies. *Clinical Neurophysiology*, 123(7):1353–1360.

Nocedal, J. (1980). Updating Quasi-Newton Matrices with Limited Storage. *Mathematics of Computation*, 35(151):773–782.

Nolte, G. and Dassios, G. (2005). Analytic expansion of the EEG lead field for realistic volume conductors. *Physics in Medicine and Biology*, 50(16):3807.

Nolte, G., Marzetti, L., and Valdes Sosa, P. (2009). Minimum Overlap Component Analysis (MOCA) of EEG/MEG data for more than two sources. *Journal of Neuroscience Methods*, 183(1):72–76.

Nolte, G., Meinecke, F. C., Ziehe, A., and Müller, K.-R. (2006). Identifying interactions in mixed and noisy complex systems. *Physical Review E*, 73:051913.

Nolte, G., Ziehe, A., Nikulin, V., Schlögl, A., Krämer, N., Brismar, T., and Müller, K.-R. (2008). Robustly Estimating the Flow Direction of Information in Complex Physical Systems. *Physical Review Letters*, 100:234101.

Nunez, P. L. and Srinivasan, R. (2005). *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, USA, 2 edition.

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

Pascual-Marqui, R. D., Michel, C. M., and Lehmann, D. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, 18:49–65.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.

Pedregosa, F., Eickenberg, M., Ciuciu, P., Thirion, B., and Gramfort, A. (2015). Data-driven HRF estimation for encoding and decoding models. *NeuroImage*, 104:209–220.

Pfurtscheller, G. and Lopes da Silva, F. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110(11):1842–1857.

Pfurtscheller, G., Neuper, C., Brunner, C., and da Silva, F. L. (2005). Beta rebound after different types of motor imagery in man. *Neuroscience Letters*, 378(3):156–159.

Pfurtscheller, G., Stancák Jr., A., and Neuper, C. (1996). Post-movement beta synchronization. A correlate of an idling motor area? *Electroencephalography and Clinical Neurophysiology*, 98(4):281–293.

Picton, T. W., John, M. S., Dimitrijevic, A., and Purcell, D. (2003). Human auditory steady-state responses. *International Journal of Audiology*, 42(4):177–219.

Plourde, G., Stapells, D. R., and Picton, T. W. (1991). The human auditory steady-state evoked potentials. *Acta Oto-laryngologica*, 491(8):153–160.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(2):676–682.

Raichle, M. E. and Snyder, A. Z. (2007). A default mode of brain function: a brief history of an evolving idea. *Neuroimage*, 37(4):1083–1090.

Ramírez, D., Schreier, P. J., Vía, J., and Nikulin, V. V. (2013). Power-CCA: Maximizing the correlation coefficient between the power of projections. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Ramoser, H., Müller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering*, 8(4):441–446.

Regan, D. (1989). *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine*. Elsevier, New York.

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

Ritter, P. and Villringer, A. (2006). Simultaneous EEG–fMRI. *Neuroscience & Biobehavioral Reviews*, 30(6):823–838.

Rodriguez, R., Picton, T., Linden, D., Hamel, G., and Laframboise, G. (1986). Human auditory steady state responses: effects of intensity and frequency. *Ear and hearing*, 7(5):300–313.

Roebroeck, A., Formisano, E., and Goebel, R. (2005). Mapping directed influence over the brain using granger causality and fMRI. *NeuroImage*, 25(1):230–242.

Romei, V., Brodbeck, V., Michel, C., Amedi, A., Pascual-Leone, A., and Thut, G. (2008). Spontaneous fluctuations in posterior alpha-band EEG activity reflect variability in excitability of human visual areas. *Cerebral Cortex*, 18(9):2010–2018.

Sänger, J., Müller, V., and Lindenberger, U. (2013). Directionality in hyperbrain networks discriminates between leaders and followers in guitar duets. *Frontiers in Human Neuroscience*, 7.

Sauseng, P., Klimesch, W., Gerloff, C., and Hummel, F. C. (2009). Spontaneous locally restricted EEG alpha activity determines cortical excitability in the motor cortex. *Neuropsychologia*, 47(1):284–288.

Scanziani, M. and Häusser, M. (2009). Electrophysiology in the age of light. *Nature*, 461(7266):930–939.

Schmidt, R. O. (1986). Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 43(3):276–280.

Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319.

Schreuder, M., Riccio, A., Risetti, M., Dähne, S., Ramsey, A., Williamson, J., Mattia, D., and Tangermann, M. (2013). User-centered design in BCI - a case study. *Artificial Intelligence in Medicine*, 59(2):71–80.

Schultze-Kraft*, M., Dähne*, S., Blankertz, B., and Curio, G. (2015). Supervised and unsupervised classification of visuomotor workload states with EEG. in preparation.

Sirotin, Y. B., Hillman, E. M., Bordier, C., and Das, A. (2009). Spatiotemporal precision and hemodynamic mechanism of optical point spreads in alert primates. *Proceedings of the National Academy of Sciences*, 106(43):18390–18395.

Sturm, I., Dähne, S., Blankertz, B., and Curio, G. (2015a). Multi-variate EEG analysis as a novel tool to examine brain responses to naturalistic music stimuli. *PLOS ONE*. in revision.

Sturm, I., Treder, M., Miklody, D., Purwins, H., Dähne, S., Blankertz, B., and Curio, G. (2015b). The polyphonic brain: Extracting the neural representation of tone onsets for separate voices of polyphonic music using multivariate EEG analysis. *Psychomusicology*. accepted.

Suk, H.-I., Fazli, S., Mehnert, J., Müller, K.-R., and Lee, S.-W. (2014). Predicting bci subject performance using probabilistic spatio-temporal filters. *PloS one*, 9(2):e87056.

Sun, L., Ji, S., Yu, S., and Ye, J. (2009). On the equivalence between canonical correlation analysis and orthonormalized partial least squares. In *International Joint Conference on Artifical Intelligence (IJCAI)*.

Taulu, S. and Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine and Biology*, 51(7):1759.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1):77–94.

Thorniley, M., Livera, L., Wickramasinghe, Y., Spencer, S., and Rolfe, P. (1990). The non-invasive monitoring of cerebral tissue oxygenation. *Advances in Experimental Medicine and Biology*, 277:323.

Thut, G., Nietzel, A., Brandt, S. A., and Pascual-Leone, A. (2006). Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *The Journal of Neuroscience*, 26(37):9494–9502.

Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. Winston Washington, DC.

Tomioka, R. and Müller, K. R. (2010). A regularized discriminative framework for EEG analysis with application to brain-computer interface. *NeuroImage*, 49:415–432.

Vapnik, V. (2000). *The nature of statistical learning theory*. springer.

Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V., and Thirion, B. (2011). Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In *Information Processing in Medical Imaging*, pages 562–573. Springer.

Varoquaux, G., Sadaghiani, S., Pinel, P., Kleinschmidt, A., Poline, J., and Thirion, B. (2010). A group model for stable multi-subject ICA on fMRI datasets. *NeuroImage*, 51(1):288–299.

Venthur, B., Dähne, S., Höhne, J., Heller, H., and Blankertz, B. (2015). Wyrm: A brain-computer interface toolbox in Python. *Journal of Neuroinformatics*, pages 1–16.

Venthur, B., Scholler, S., Williamson, J., Dähne, S., Treder, M. S., Kramarek, M. T., Müller, K.-R., and Blankertz, B. (2010). Pyff – a pythonic framework for feedback applications and stimulus presentation in neuroscience. *Frontiers in Neuroscience*, 4:179.

Vulliemoz, S., Carmichael, D. W., Rosenkranz, K., Diehl, B., Rodionov, R., Walker, M. C., McEvoy, A. W., and Lemieux, L. (2011). Simultaneous intracranial EEG and fMRI of interictal epileptic discharges in humans. *NeuroImage*, 54(1):182–190.

Wakeman, D. G. and Henson, R. N. (2015). A multi-subject, multi-modal human neuroimaging dataset. *Scientific Data*, 2.

Weiller, D., Märtin, R., Dähne, S., Engel, A. K., and König, P. (2010). Involving motor capabilities in the formation of sensory space representations. *PLOS ONE*, 5(4):e10377.

Wichmann, T. and Delong, M. R. (2011). Deep-Brain stimulation for basal ganglia disorders. *Basal Ganglia*, 1(2):65–77.

Winkler, I., Haufe, S., Porbadnigk, A., Müller, K.-R., and Dähne, S. (2015). Identifying Granger causal relationships between neural power dynamics and variables of interest. *NeuroImage*, 111:489–504.

Wong, D.-O., Hwang, H.-J., Dähne, S., Müller, K.-R., and Lee, S.-W. (2015). Effect of higher frequency on steady state visual evoked potential based brain-computer interface. *Journal of Neural Engineering*. in revision.

Wray, S., Cope, M., Delpy, D. T., Wyatt, J. S., and Reynolds, E. O. R. (1988). Characterization of the near infrared absorption spectra of cytochrome aa_3 and haemoglobin for the non-invasive monitoring of cerebral oxygenation. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 933(1):184–192.

Wyler, A. R., Ojemann, G. A., Lettich, E., and Ward Jr, A. A. (1984). Subdural strip electrodes for localizing epileptogenic foci. *Journal of Neurosurgery*, 60(6):1195–1200.

Ziehe, A. and Müller, K.-R. (1998). TDSEP – an efficient algorithm for blind separation using time structure. In Niklasson, L., Bodén, M., and Ziemke, T., editors, *Proc. of the 8th International Conference on Artificial Neural Networks, ICANN'98*, Perspectives in Neural Computing, pages 675 – 680, Berlin. Springer Verlag.