UNIVERSITY OF CALIFORNIA SAN DIEGO

Towards electromagnetic source imaging methods for developing brain-computer interface neurotherapeutics

A dissertation submitted in partial satisfaction of the
requirements for the degree of Doctor of Philosophy

in

Electrical Engineering
(Intelligent Systems, Robotics, and Control)

by

Alejandro Ojeda

Committee in charge:

Kenneth Kreutz-Delgado, Chair
Jyoti Mishra Ramanathan, Co-Chair
Scott Makeig
Terrence J. Sejnowski
Bhaskar Rao
Nuno M. Vasconcelos

2019

The Dissertation of Alejandro Ojeda is approved and is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California San Diego

2019

DEDICATION

To Susana and baby Sebastian with all my love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge my family for their unconditional support throughout this journey. Mom, dad, siblings, grandma, and grandpa (who's writing a book at his 83's), I love you all.

I want to thank my advisor Professor Ken Kreutz-Delgado for his support throughout grad school and for believing in me from the start. I learned from Ken the foundations of modern data science from the very first principles to the state of the art. At the time it seemed like a monumental effort, and it was, but he managed to teach it all in two quarters. In his classes, he uses a storytelling style that demonstrates that math doesn't need to be boring or excessively difficult, but it is just like another language that you can pick up if you try. While writing papers together, I learned from Ken the importance of being rigorous and to think deeply about every statement, these teachings will continue with me as I navigate the academic path.

I want to thank my co-advisor Professor Jyoti Mishra for offering me a plethora of interesting biologically-relevant problems to which I could apply my methods, as well as sharing her experience in translational research. I also thank Professor Mishra for pushing me out of my comfort zone, encouraging, and helping me to write grants that have some chance of getting funded. In my experience, grant writing is a daunting task for most engineering students. I am very grateful that I could learn a few things about this art from her and I hope that by getting into the habit of doing it, I will get better at it eventually.

I want to thank Dr. Scott Makeig and Professors Nuno Vasconcelos, Bhaskar Rao, and Terry Sejnowski for taking the time to participate in my doctoral committee. I am especially grateful to Scott, who brought me to San Diego to work with him and so many bright people at the Swartz Center for Computational Neuroscience (SCCN) a few years ago and I haven't been able to leave the city to this day.

I want to acknowledge my colleagues from the SCCN, especially Ramon Martinez-Cancino, who suggested me to use information theoretic metrics to characterize brain

dynamics, which turned out to be a very fruitful approach. I thank my colleagues from the Neural Engineering and Translation Labs, your intellectual curiosity is unbounded, and I am very grateful that I share lab space with you all every day. Also, I would like to acknowledge my teachers, colleagues, and friends from the Cuban Neuroscience Center from whom I first learned the foundations of brain imaging.

Chapter 2, in part, has been submitted for publication, and it is available as a bioRxiv preprint in Ojeda, A., Klug, M., Kreutz-Delgado, K., Gramann, K., & Mishra, J. (2019). A Bayesian framework for unifying data cleaning, source separation and imaging of electroencephalographic signals. https://doi.org/10.1101/559450. I was the primary investigator and author of this paper.

Chapter 3, in part, is a reprint of the material as it appears in Ojeda, A., Kreutz-Delgado, K., & Mullen, T. (2018). Fast and robust Block-Sparse Bayesian learning for EEG source imaging. NeuroImage, 174(8), 449-462. https://doi.org/10.1016/j.neuroimage.2018.03.048. I was the primary investigator and author of this paper.

Chapter 4 is currently being prepared for submission for publication of the material. I am the primary investigator and author of this material along with co-authors Nathalie Buscher, Pragathi Priyadharsini Balasubramani, Vojislav Maric, Dhakshin Ramanathan, and Jyoti Mishra.

Chapter 5 is currently being prepared for submission for publication of the material. I am the primary investigator and author of this material along with co-authors Margot Wagner and Jyoti Mishra.

VITA

2006	Bachelor of Engineering in Automation and Control, Polytechnic University of Havana (a.k.a. CUJAE), Cuba
2009	Master of Science in Neurophysics and Neuroengineering, University of Havana, Cuba
2019	Doctor of Philosophy, University of California, San Diego
2006–2010	Research Scientist, Cuban Neuroscience Center, Havana, Cuba
2011–2013	Programmer Analyst, Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California, San Diego
2013–2017	Research Scientist and Co-Founder, Syntrogi, Inc., San Diego, California
2017-2019	Graduate Student Researcher, Neural Engineering and Translation Labs, University of California, San Diego

PUBLICATIONS

Ojeda, A. et al. (2019) A Bayesian framework for unifying data cleaning, source separation and imaging of electroencephalographic signals. doi: 10.1101/559450.

Ojeda, A., Kreutz-Delgado, K. and Mullen, T. (2018) Fast and robust Block-Sparse Bayesian learning for EEG source imaging. NeuroImage, 174(8), pp. 449-462. doi: 10.1016/j.neuroimage.2018.03.048.

Nonoda, Y., Miyakoshi, M., **Ojeda**, A., Makeig, S., Juhasz, C., Sood, S., & Asano, E. (2016). Interictal high-frequency oscillations generated by seizure onset and eloquent areas may be differentially coupled with different slow waves. Clinical Neurophysiology, 127(6), 2489-2499.

Mullen, T. R., Kothe, C. A. E., Chi, M., **Ojeda**, A., Kerth, T., Makeig, S., Cauwenberghs, G. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. IEEE Transactions on Biomedical Engineering, 62(11), pp. 2553-2567. doi: 10.1109/TBME.2015.2481482.

Iversen, J., **Ojeda**, A., Mullen, T., Plank, M., Snider, J., Cauwenberghs, G., Poizner, H. (2014). Causal analysis of cortical networks involved in reaching to spatial targets. Conference proceedings of IEEE Engineering in Medicine and Biology Society, pp: 4399-4402, doi:10.1109/EMBC.2014.6944599.

T., Khalil, A., Ward, T., Iversen, J., Leslie, G., Warp, R., Whitman, M., Minces, V., McCoy, A., **Ojeda**, A., Bigdely-Shamlo, N., Chi, M., and Rosenboom, D. (2015). MindMusic: Playful and Social Installations at the Interface Between Music and the Brain. In A. Nijholt (Ed.), More Playful User Interfaces. Interfaces that Invite Social and Physical

Interaction. Singapore: Springer Singapore.

Leslie, G., **Ojeda**, A., & Makeig, S. (2014). Measuring musical engagement using expressive movement and EEG brain dynamics. Psychomusicology: Music, Mind, and Brain, 24(1), 75-91.

Ojeda, A., Bigdely-Shamlo, N., & Makeig, S. (2014). MoBILAB: an open source toolbox for analysis and visualization of mobile brain/body imaging data. Frontiers in Human Neuroscience, 8, 121.

Leslie, G., **Ojeda**, A., & Makeig, S. (2013). Towards an Affective Brain-Computer Interface Monitoring Musical Engagement. In 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (pp. 871-875). IEEE.

Valdes-Hernandez, P. A., **Ojeda**, A., Martinez-Montes, E., Lage-Castellanos, A., Virues-Alba, T., Valdes-Urrutia, L., & Valdes-Sosa, P. A. (2010). White matter architecture rather than cortical surface area correlates with the EEG alpha rhythm. NeuroImage, 49(3), 2328-39.

Valdes-Hernandez, P. A., von Ellenrieder, N., **Ojeda**, A., Kochen, S., Aleman-Gomez, Y., Muravchik, C., & Valdes-Sosa, P. A. (2009). Approximate average head models for EEG source imaging. Journal of Neuroscience Methods, 185(1), 125-132.

ABSTRACT OF THE DISSERTATION

Towards electromagnetic source imaging methods for developing brain-computer interface neurotherapeutics

by

Alejandro Ojeda

Doctor of Philosophy in Electrical Engineering
(Intelligent Systems, Robotics, and Control)

University of California San Diego, 2019

Kenneth Kreutz-Delgado, Chair
Jyoti Mishra Ramanathan, Co-Chair

Despite several decades of research, most mental health treatments are based on pharmacological manipulations that globally affect the nervous system. Such treatments often lead to undesired side effects and short term symptomatic relief. The difficulty of diagnosing and treating mental health illnesses stems from the overwhelming complexity of the brain and is exacerbated by the fact that our ability to probe, simultaneously, the activity of dynamic and distributed brain networks is limited. In this dissertation, I propose an alternative way to tackle the mental health problem by using high-resolution imaging-

based brain-computer interface (BCI) neurotechnology. I focus on new neuroimaging technology that allows us to monitor the electrical activity of cortical networks at low-cost and high spatiotemporal resolution using noninvasive electroencephalographic (EEG) measurements. This technology will serve as the "neural decoder" component of yet to come imaging-based closed-loop systems that can effectively restore impaired cognition. The decoder allows a BCIs to dynamically probe specific cognitive abilities of the subject in search for signatures of circuit dysfunctions. Then, various types of feedback can be designed to induce the engagement of neural populations that can compensate for the detected aberrant neuronal activity.

In this dissertation, first, I develop the mathematical framework to efficiently map scalp EEG responses back into the cortical space, and by doing so, I show that the biological mechanisms responsible for the neurocognitive processes of interest are easy to study. Of theoretical and practical relevance, I demonstrate that this framework successfully unifies three of the most common problems in EEG analysis: data cleaning, source separation, and imaging. Then, I develop the algorithmic and software machinery necessary to implement high-resolution imaging-based BCIs. Finally, I analyze data from healthy adults performing a self-paced unconstrained schoolwork-like computerized task and show that within the proposed framework, I can identify brain network correlates of attention switches at a millisecond time scale. Since attention-related dysfunctions are linked to several psychiatric disorders, these results represent a step forward towards developing BCI interventions to treat several mental health illnesses.

Chapter 1

Introduction

1.1 The mental health problem

A mental illness can be defined as a mental, behavioral, or emotional disorder that can result in mild, moderate, or severe functional impairments that interfere with life activities. These conditions include depression, anxiety and bipolar disorders, attention-deficit/hyperactivity disorder, autism, schizophrenia, among others. Despite substantial advances in drug therapies, according to the National Institute of Mental Health (NIMH), nearly one in five adults in the U.S. lives with a mental health illness[1], that was 46.6 million in 2017.

The difficulty of diagnosing and treating these conditions stems from the facts that their genetic, neurological, and environmental causes remain poorly understood. Mental illnesses are likely to be the result of dysfunctional neural circuits [Akil et al., 2010]. Developmental, structural, and functional defects in these circuits can lead to a plethora of symptoms. And since thousands of genes are involved in the regulation of neural development and function, abnormalities in one or more of them can lead to complex neuropsychiatric phenotypes. In addition, cellular metabolic dysfunctions and brain lesions can disturb brain circuits in complicated ways, thereby resulting in ambiguous clinical manifestations. This overwhelming complexity is exacerbated by the fact that our ability

[1] https://www.nimh.nih.gov/health/statistics/mental-illness.shtml

to monitoring the functional activity of dynamic brain networks is limited.

In this dissertation, I focus on the development of neuroimaging technology that allows us to monitor the electrical activity of cortical networks in a noninvasive way, with low-cost, and high spatiotemporal resolution. This technology can serve as the "neural decoder" component of yet to come imaging-based closed-loop brain-computer interface (BCI) systems in which we can dynamically probe specific cognitive abilities in search for signatures of circuit dysfunctions while giving feedback targeted to inducing the engagement of neural populations that can compensate the aberrant neuronal activity (see Fig. 1.1). Although the feedback component of such a system is beyond the scope of this dissertation, in Chapter 4 I develop a library that can aid in its implementation.

1.2 EEG source imaging BCI in mental health

We can define a BCI as a bidirectional communication channel between a brain and a computer with the purpose of compensating dysfunctional neuronal activity, rehabilitating motor skills, or enhancing a cognitive ability. Likewise, a closed-loop BCI can be defined as a BCI that uses stimulation and feedback loops to adapt its communication channel to ongoing changes in brain dynamics, task goals, and environmental signals. A promising application of closed-loop BCI technology to clinical neuroscience is the development of personalized computerized therapies that can target specific neural circuit dysfunctions leading to psychiatric disorders.

The electroencephalogram (EEG) is appealing for developing BCI neurotherapeutics because it is noninvasive, has an excellent temporal resolution, and is now available as low-cost mobile technology. Traditional EEG-based BCIs are implemented using features (neuromarkers) derived from the sensor data collected on the scalp. However, since the electrical activity of EEG sources is mixed by several layers of tissue while propagating from the cortex to the scalp, sensor-space features constitute a loose reflection of the underlying

brain processes of interest and are often confounded by other non-neural (ocular/muscular) responses unrelated to the targeted cognitive state. Furthermore, even if we can decode cognitive and behavioral states form scalp-space features with high accuracy, that doesn't help us to further our understanding of the brain network mechanisms responsible for mental illnesses. Thus, despite impressive advances in signal processing and machine learning algorithms, the use of BCI in mental health remains limited.

The central theme of this dissertation is the development of the mathematical framework that allows us to map brain responses measured by EEG back into a space where the neurobiological mechanisms responsible for the cognitive behaviors of interest are easy to study. In contrast to the common black-box approach followed by most machine learning practitioners, here we want to learn about neural circuit dysfunctions that lead to mental illnesses, so we need to be able to interpret the parameters of our models in a neurobiological context. Therefore, imaging-based BCIs have the potential to fulfill two important objectives of modern clinical neuroscience: 1) to characterize the brain network dysfunctions that lead to psychiatric illnesses and 2) to develop non-invasive circuit-specific personalized neurocognitive interventions.

To support closed-loop BCI applications such as the one depicted in Fig. 1.1, EEG signal processing algorithms need to be fast and robust to artifacts. Although the methods presented here can be used to analyze data offline, in which case processing speed is not a concern, in online systems, we usually monitor and perturb brain dynamics that evolve at a millisecond time scale; thus we need processing algorithms that can operate at that speed. Furthermore, to design neurotechnology that can function successfully in the real world (i.e., outside of a tightly controlled lab environment and without user intervention), we need to obtain data representations that adapt to the statistics of the sensor noise, minimizing the effect of biological and environmental artifacts, while keeping up with the speed of the measurements. I show that the framework proposed here unifies three of the most common problems in EEG analysis: data cleaning, source separation, and source

imaging. The solution of these dissimilar problems can be understood as the optimization of a quantity known as the Bayesian evidence of a biologically inspired model of how cortical sources give rise to the noisy EEG voltages measured on the scalp. Furthermore, I show that his framework yields an implementation that can achieve real-time performance.

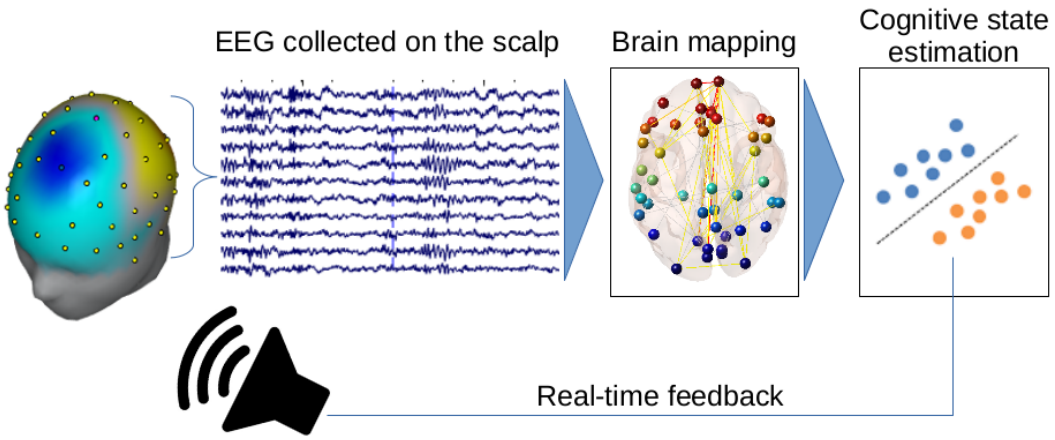


Figure 1.1. Schematic of the proposed closed-loop neurotechnology. EEG time series are collected while the subject engages naturally in daily life activities. A processing pipeline estimates the inverse mapping from signals in the sensor space onto the cortex minimizing the effects of artifacts without user intervention. Neural features, such as source power and connectivity can be calculated and passed onto a machine learning module to perform a cognitive state estimation. Depending on the cognitive state, feedback may be delivered to steer the mental state of the subject towards the desired target.

1.3 Attention regulation in mental illness

Since developing closed-loop neurotechnology that can monitor, restore, and enhance general cognitive abilities may be a career-long endeavor, as a proof of principle, here I study the brain network mechanisms that may be involved in the switch of attention in an unconstrained self-paced schoolwork computerized task. Since attention-related brain network dysfunctions have been observed in several psychiatric disorders [Arnsten and Rubia, 2012; Millan et al., 2012], including attention-deficit/hyperactivity disorder, post-traumatic stress disorder, schizophrenia, and even mood disorders such as depression

and anxiety, I use this example to demonstrate the potential of the proposed imaging framework to unveil the neural mechanisms of mental illnesses and develop effective BCI neurotherapeutics.

1.4 How to read this document and outline

The rest of the dissertation is organized as follows. In Chapter 2, I develop a unifying Bayesian framework for EEG data cleaning, source separation, and imaging; this is the methodological core of the dissertation. Chapters 3-4 build up the algorithmic and software machinery to implement imaging-based BCIs. In Chapter 3, I redesign parts of the block-sparse Bayesian learning algorithm to yield sensor noise adaptation and EEG source estimates in real-time; although the algorithm is developed assuming artifact-free EEG data, it can be recast to invert the probabilistic generative model proposed in Chapter 2. In Chapter 4, I develop SimBSI, a graphical library for the rapid prototyping of closed-loop BCIs using the brain imaging methods developed in earlier chapters. In Chapter 5 I apply the framework of Chapter 2 to the study of the cortical networks involved in the switch of attention of healthy adult subjects during an unconstrained schoolwork-like computerized task.

Each chapter is as auto-contained as possible, so in principle, they can be read in any order. The left panel of Fig. 1.2 shows the default order, which may be more suitable for readers with a background in neuroimaging, signal processing, and data science. This order goes from the mathematical framework to algorithm and software development to a cognitive state decoding application. The order displayed on the right panel may be more appropriate for readers with a background in cognitive sciences or mental health, as it starts with the application and software tools for BCI and then progresses onto the mathematical framework.

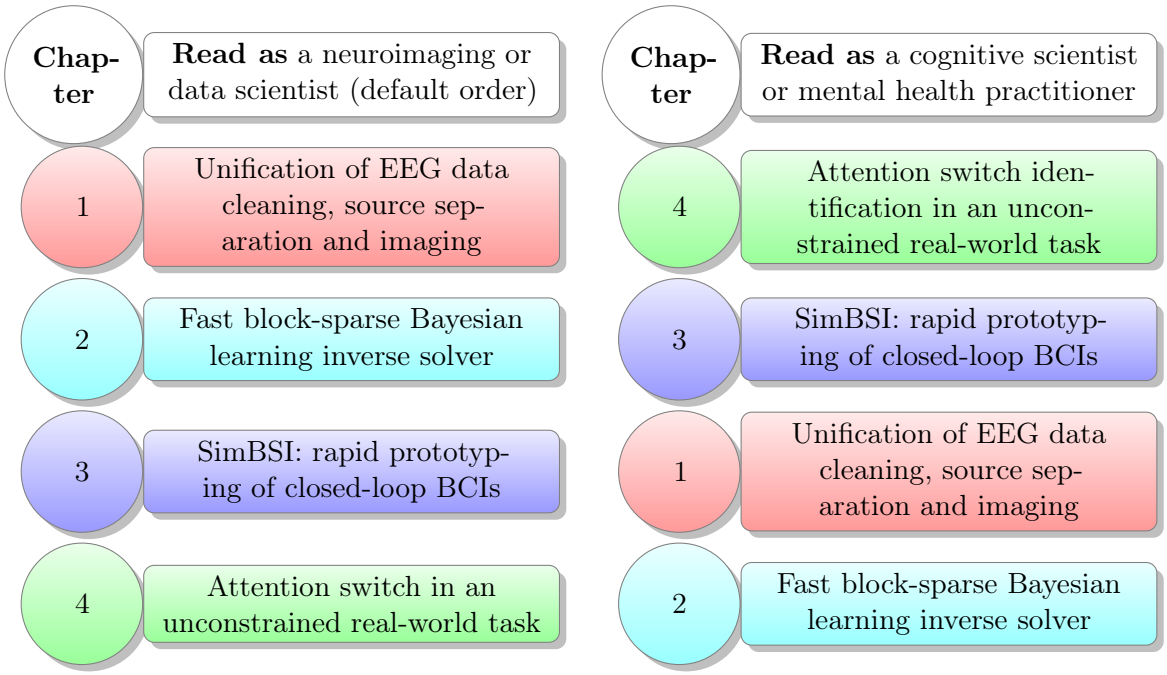


Figure 1.2. Recommended reading order of this dissertation depending on reader's background and interest. At a glance, the dissertation is organized as follows: Chapter 2 is the theoretical framework, Chapters 3-4 build up the algorithmic and software machinery to analyze imaging data and implement closed-loop BCIs, and we finalize in Chapter 5 with an application of EEG source imaging to the study of attention switches in a real-world computerized task.

1.5 Notation

Unless otherwise stated, throughout this document I use the following notation. Bold lowercase and uppercase characters or symbols denote column-wise vectors and matrices respectively. $\mathbf{I}_N$ represents a $N \times N$ identity matrix. $\hat{\mathbf{x}}$ is the estimate of parameter $\mathbf{x}$. The character N is used to denote the length of an array along some dimension, e.g., if $\mathbf{x} \in \mathbb{R}^{N_x}$, then N_x denotes the length of vector $\mathbf{x}$. Lastly, $\|\mathbf{A}\|_{\mathcal{F}}^2$ denotes the square of the Frobenius norm of matrix $\mathbf{A}$, $\text{trace}(\mathbf{A}^T\mathbf{A})$.

Chapter 2

A Bayesian framework for unifying data cleaning, source separation and imaging of electroencephalographic signals

Abstract

Electroencephalographic (EEG) source imaging depends upon sophisticated signal processing algorithms for data cleaning, source separation, and localization. Typically, these problems are addressed by independent heuristics, limiting the use of EEG images on a variety of applications. Here, we propose a unifying parametric empirical Bayes framework in which these dissimilar problems can be solved using a single algorithm (PEB+). We use sparsity constraints to adaptively segregate brain sources into maximally independent components with known anatomical support, while minimally overlapping artifactual activity. Of theoretical relevance, we demonstrate the connections between Infomax ICA and our framework. On real data, we show that PEB+ outperforms Infomax for source separation on short time-scales and, unlike the popular ASR algorithm, it can reduce artifacts without significantly distorting clean epochs. Finally, we analyze mobile brain/body imaging data to characterize the brain dynamics supporting heading computation during full-body rotations, replicating the main findings of previous experimental

literature.

2.1 Introduction

The electroencephalogram (EEG) is a noninvasive functional brain imaging modality that allows the study of brain electrical activity with excellent temporal resolution. Compared to other noninvasive imaging modalities such as fMRI, PET, SPECT, and MEG, EEG acquisition can be mobile and more affordable [Mcdowell et al., 2013; Mehta and Parasuraman, 2013], allowing the widespread study of human cognition and behavior under more ecologically valid experimental conditions [Makeig et al., 2009]. Imaging cognitive processes while participants engage naturally with their environment (natural cognition in action [Gramann et al., 2014]) has potential for developing a new generation of applications in brain-computer interfaces (BCI), mental health, rehabilitation, and neuroergonomics [Jungnickel and Gramann, 2016; Mishra and Gazzaley, 2014; Wagner et al., 2016]. However, despite impressive methodological advances in the estimation of the electrical activity of the cortex from EEG voltages recorded on the scalp, a number of practical and theoretical issues remain unsolved.

Imaging EEG source activity (also known as electromagnetic source imaging or ESI) is challenging for several reasons. First, since many configurations of currents in the brain can elicit the same EEG scalp topography [Michel and Murray, 2012], it entails solving an ill-posed inverse problem [Lopes da Silva, 2013]. Second, the EEG signal is often contaminated by artifacts of non-brain origin such as electrooculographic (EOG) and electromyographic (EMG) activity that need to be identified and removed. Third, there is evidence that large-scale brain responses measured by EEG are generated by underlying cortical dynamics that evolve over time and can exhibit nonlinear features [Breakspear, 2017; Khambhati et al., 2018], thereby rendering the simplifying assumptions of linearity and stationarity used by most inverse methods hard to justify. These problems are usually

addressed separately using a variety of heuristics, making it difficult to systematize a methodology for obtaining biologically plausible EEG source estimates in the presence of artifacts and nonlinear and nonstationary dynamics. The objective of this chapter is to develop a unifying Bayesian framework in which these, apparently dissimilar, problems can be understood and solved in a principled manner using a single algorithm.

To cope with the ill-posed nature of the inverse problem and ensure functional images with biological relevance, several inverse algorithms have been proposed that seek to estimate EEG sources subject to neurophysiologically reasonable spatial [Baillet et al., 2001; Friston et al., 2008; Haufe et al., 2011; Pascual-Marqui et al., 2002; Trujillo-Barreto et al., 2004], spatiotemporal [Martínez-Vargas et al., 2015; Trujillo-Barreto et al., 2008; Valdés-Sosa et al., 2009], and frequency-domain [Gramfort et al., 2013] constraints, just to mention a few examples. These approaches can work relatively well when the EEG samples are corrupted by Gaussian noise and the signal to noise ratio (SNR) is high. In practice, however, raw EEG data are affected by many other types of noise such as interference from the 50/60 Hz AC line, pseudo-random muscle activity, and mechanically induced artifacts, among others. Thus, before source estimation, non-Gaussian artifacts need to be removed from the data.

There is a plethora of methods for dealing with artifacts corrupting the EEG signal [Islam et al., 2016; Mannan et al., 2018]. Popular approaches used in real-time BCI applications are based on adaptive noise cancellation [Kilicarslan et al., 2016] or Artifact Subspace Removal (ASR) [Mullen et al., 2015] algorithms. The former has the inconvenience that an additional channel recording purely artifactual activity (i.e., EOG or EMG activity not admixed with EEG) needs to be provided, while the latter rests on the assumption that the statistics of data and artifacts stay the same after an initial calibration phase. In studies where the data can be analyzed offline, artifactual components can be largely removed using Independent Component Analysis (ICA) [Jung et al., 2000]. ICA-based cleaning, however, has the drawback that non-brain components need to be identified

for removal, which is usually done manually based on the practitioner's experience.

ICA is a special case of blind source separation (BSS) method [Cichocki and Amari, 2002] that can be used to linearly decompose EEG data into components that are maximally statistically independent. ICA has been used to analyze event-related potentials (ERP) under the assumptions that during the task 1) the decomposition is stationary and 2) that brain components can be modeled as a predefined number of dipolar point processes with fixed spatial location and orientation [Makeig and Onton, 2011]. The stationarity assumption can be relaxed using a mixture of ICA models [Palmer et al., 2011] while the selection of brain scalp projections is typically done either manually or automatically based on the residual variance afforded by a dipole fitting algorithm. The practical use of ICA has been limited by its computational cost and the need for user intervention. Only recently, a real-time recursive ICA algorithm has been proposed [Hsu et al., 2016], as well as a number of automatic methods for minimizing the subjectivity of manual component selection [Pion-Tonachini et al., 2017; Radüntz et al., 2017; Tamburro et al., 2018]. Despite these advances, turning ICA into a brain imaging modality requires that after source separation, we solve the inverse problem of localizing the set of identified brain components into the cortical space.

One way of estimating EEG sources subject to multiple assumptions (constraints) in a principled manner is to use the framework of parametric empirical Bayes (PEB) [Casella, 1985; Morris, 1983]. In this framework, constraints are used to furnish prior probability density functions (pdfs). Empirical Bayes methods use data to infer the parameters controlling the priors (hyperparameters), such that those assumptions that are not supported by the data can be automatically discarded without user intervention. Here we use priors to "encourage" source images to belong to a functional space with biological relevance, but the exact form of those priors is determined by the data (empirically). In the context of sparsity-inducing priors, PEB is sometimes referred to as Sparse Bayesian Learning (SBL) [Tipping, 2001]. The PEB framework has been applied extensively to

brain imaging for inverting hierarchical models of fMRI and PET [Friston et al., 2002] as well as EEG responses [Friston et al., 2008; Henson et al., 2011].

In this chapter, within the PEB framework, we propose a probabilistic generative model (PGM) of how the raw EEG time series arises from the superposition of the brain and non-brain (artifact) source activity. We use an anatomical brain atlas to parameterize the source prior such that we induce sparsity in the number of cortical regions active at any given time (this is known in the SBL literature as block-sparsity [Zhang and Rao, 2013]). Then, our model can be inverted using an efficient block SBL algorithm recently proposed by Ojeda et al. [2018]. Henceforth, we refer to this new approach as PEB+ (PEB with the addition of artifact modeling). Our main contribution is that, by explicitly modeling non-brain sources within the PEB framework, we can unify three of the most common problems in EEG analysis: data cleaning, source separation, and source imaging. In addition, we show that the PEB+ algorithm has online adaptation, thereby allowing it to capture transient brain dynamics. On the theoretical side, we point out the connections between distributed source imaging and ICA, two popular approaches that are often perceived to be at odds with one another.

The rest of the chapter is organized into two main sections concerned with methods and results respectively. **Methods:** In Section 2.2.1 we propose the augmented PGM of the EEG taking into account the contribution of artifact sources and motivate all our modeling assumptions. In Sections 2.2.3-2.2.4 we point out the connections between PEB+ and ICA. **Results:** In Section 2.3.1 we explain how to construct an empirical dictionary of artifact scalp projections. In Sections 2.3.2, 2.3.3, and 2.3.4 we investigate the respective source imaging, separation, and data cleaning capabilities of the PEB+ algorithm.

2.2 Methods

It has been shown that popular source estimation algorithms used in ESI such as weighted minimum l_2-norm [Baillet et al., 2001], FOCUSS [Cotter et al., 2005; Gorodnitsky and Rao, 1997], minimum current estimation [Huang et al., 2006], sLORETA [Pascual-Marqui et al., 2002], beamforming [Van Veen et al., 1997], variational Bayes [Friston et al., 2008], and others can be expressed in a unifying Bayesian framework [Wipf and Nagarajan, 2009]. We extend this framework by explicitly modeling non-brain artifact sources.

2.2.1 Augmented probabilistic generative model

In source imaging, the neural activity is often referred to as the primary current density (PCD) [Baillet et al., 2001] and it is defined on a grid of known cortical locations (the source space). Typically, a vector of N_y EEG measurements at sample k, $\mathbf{y}_k \in \mathbb{R}^{N_y}$, relates to N_g PCD values, $\mathbf{g}_k \in \mathbb{R}^{N_g}$, through the following linear equation [Dale and Sereno, 1993],

$$\mathbf{y}_k = \mathbf{L}\mathbf{g}_k + \mathbf{e}_k, \quad k = 1, \ldots, N \tag{2.1}$$

where $\mathbf{e}_k \in \mathbb{R}^{N_y}$ represents the measurement noise vector. The PCD is projected to the sensor space through the lead field matrix $\mathbf{L} = [\mathbf{l}_1, \ldots, \mathbf{l}_{N_g}] \in \mathbb{R}^{N_y \times N_g}$ ($N_y \ll N_g$) where each column $\mathbf{l}_i$ denotes the scalp projection of the ith unitary current dipole with fixed orientation within the source space. When dipole orientations are considered, then $\mathbf{L} \in \mathbb{R}^{N_y \times 3N_g}$ and we determine a source vector $\mathbf{g}_k \in \mathbb{R}^{3N_g}$. The lead field matrix is usually precomputed for a given electrical model of the head derived from a subject-specific MRI [Hallez et al., 2007]. Alternatively, if an individual MRI is not available, an approximated lead field matrix obtained from a high-resolution template can be used [Huang et al., 2016]. Then, the inverse problem of the EEG can be stated as the estimation of a source configuration $\hat{\mathbf{g}}_k$ that is likely to produce the scalp topography $\mathbf{y}_k$.

In the generative model presented above, the noise term $\mathbf{e}_k$ is assumed to be

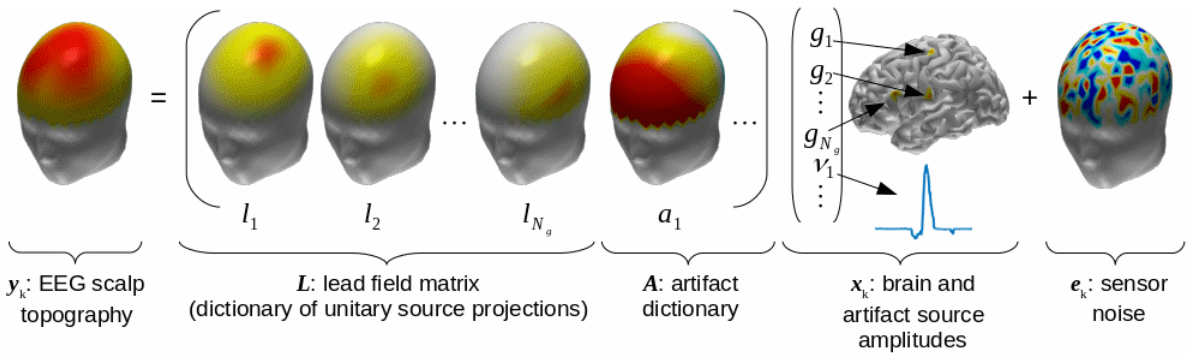


Figure 2.1. Proposed augmented generative model of the EEG. The model postulates that the EEG scalp topography $\mathbf{y_k}$ arises from the linear superposition of brain $g_{i,k}$ and artifact $\nu_{j,k}$ components weighted by their respective scalp projections $\mathbf{l_i}$ and $\mathbf{a_j}$, corrupted by spatially uncorrelated Gaussian noise $\mathbf{e_k}$.

Gaussian and spatially uncorrelated with variance λ. This simplification is acceptable as long as EEG topographies are not affected by non-Gaussian pseudo-random artifacts generated by eye blinks, lateral eye movements, facial and neck muscle activity, body movement, among others. Therefore, before source estimation, EEG data are usually heavily preprocessed and cleaned [Bigdely-Shamlo et al., 2015]. Since artifacts contribute linearly to the sensors, ideally, one would like to characterize their scalp projections to describe more accurately the signal acquisition. To this end, we propose the following generalization of Eq (2.1),

$$\mathbf{y}_k = \mathbf{L}\mathbf{g}_k + \mathbf{A}\boldsymbol{\nu}_k + \mathbf{e}_k \tag{2.2}$$

where $\boldsymbol{\nu} \in \mathbb{R}^{N_\nu}$ is a vector of N_ν artifact sources and $\mathbf{A} = [\mathbf{a_1}, \ldots, \mathbf{a_{N_\nu}}] \in \mathbb{R}^{\mathbf{N_y} \times \mathbf{N}_\nu}$ is a dictionary of artifact scalp projections (see Fig 2.1).

Although the entries of $\mathbf{A}$ that correspond to muscle activity may be obtained based on a detailed electromechanical model of the body [Böl et al., 2011], in most studies this approach may not be feasible due to computational and budgetary constraints. Janani et al. [2017] modelled $\mathbf{A}$ by expanding the lead field matrix to account for the contribution of putative scalp sources, which were assumed to be the generators of EMG activity. They used sLORETA to estimate brain and scalp sources simultaneously. Although this

approach was shown to be as effective as ICA-based artifact removal, it was suggested by the authors that the use of the non-sparse solver sLORETA may lead to unrealistic configurations of brain and non-brain sources. Similarly, Fujiwara et al. [2009] augmented the magnetic lead field matrix to model the scalp contribution of two current dipoles located behind the eyes and used a Bayesian approach, that has similarities with ours, to estimate brain and eye source activity from MEG data. Although successful for removing EOG activity, in their formulation, Fujiwara et al. [2009] ignored other types of artifacts that are harder to model such as those produced by muscular activity.

In this chapter, we take an empirical view inspired by the success of ICA-based artifact removal approaches. We propose constructing the dictionary $\mathbf{A}$ using a set of stereotypical artifact scalp projections such as those obtained from applying ICA to a database of EEG recordings [Bigdely-Shamlo et al., 2013a]. We then rewrite Eq (2.2) in a compact manner as follows,

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{e}_k \tag{2.3}$$

where the gain matrix is now $\mathbf{H} \triangleq [\mathbf{L}, \mathbf{A}]$ and $\mathbf{x}_k \triangleq [\mathbf{g}_k^T, \boldsymbol{\nu}_k^T]^T$ is the augmented vector of hidden (latent) brain and artifact sources (see Fig 2.1).

Note that, structurally, the standard generative model in Eq (2.1) and the augmented one in Eq (2.3) are identical. They differ however in that in Eq (2.3) we are explicitly modeling the instantaneous spatial contribution of non-brain sources to the scalp topography $\mathbf{y}_k$. Therefore, we may be able to dispense with computationally expensive preprocessing data cleaning procedures. The assumption of Gaussian measurement noise yields the following likelihood function,

$$p(\mathbf{y}_k|\mathbf{x}_k, \lambda) = N(\mathbf{y}_k|\mathbf{H}\mathbf{x}_k, \lambda\mathbf{I}_{N_y}) \tag{2.4}$$

Since Eq (2.3) does not have a unique solution, to obtain approximated source

maps with biological interpretation we introduce constraints. One way of incorporating constraints in a principled manner is to express them in the form of the prior pdf of the sources $p(\mathbf{x}_k)$. Since the neural generators of the EEG are assumed to be the electrical currents produced by distributed neural masses that become locally synchronized in space and time [Nunez and Srinivasan, 2006], here we chose a parameterized prior $p(\mathbf{x}_k)$ that induces source maps to be globally sparse (seeking to explain the observed scalp topography by a few spots of cortical activity) and locally correlated (so that we obtain spatially smooth maps as opposed to maps formed by scattered isolated sources). Artifactual sources, on the other hand, can be assumed to be uncorrelated from one another. We use a Gaussian prior to express these modeling assumptions as follows,

$$p(\mathbf{x}_k|\boldsymbol{\gamma}) = N(\mathbf{x}_k|0, \boldsymbol{\Sigma}_x) \tag{2.5}$$

where the covariance matrix $\boldsymbol{\Sigma}_x$ has a block diagonal structure [Zhang and Rao, 2013] defined as

$$\boldsymbol{\Sigma}_x = \begin{bmatrix} \boldsymbol{\Sigma}_g & \\ & \boldsymbol{\Sigma}_\nu \end{bmatrix} \tag{2.6}$$

In Eq (2.6), the brain source prior covariance is defined as

$$\boldsymbol{\Sigma}_g = \begin{bmatrix} \gamma_1 C_1 & & \\ & \ddots & \\ & & \gamma_{N_{ROI}} C_{N_{ROI}} \end{bmatrix} \tag{2.7}$$

and $\boldsymbol{\Sigma}_\nu = \text{diag}(\gamma_{N_{ROI}+1}, \ldots, \gamma_{N_{ROI}+N_\nu})$ is the covariance of artifact sources. The matrices $\mathbf{C}_i \in \mathbb{R}^{N_i \times N_i}$ encode the intra-group brain source covariances and are precomputed based on source distance taking into account the local folding of the cortex. $\boldsymbol{\gamma} \in \mathbb{R}^{N_{ROI}+N_\nu}$ denotes a nonnegative scale vector that encodes the sparsity profile of the group of sources. Here we define $N_{ROI} = 148$ groups based on anatomical regions of interest (ROI) obtained

from the Destrieux cortical atlas [Destrieux et al., 2010]. The parameterization of $\mathbf{\Sigma}_g$ based on a cortical atlas is described greater detail in Chapter 3. We note that although other atlases could be used within the framework outlined in this section (e.g. the popular Desikan-Killiany 68-ROI cortical atlas [Desikan et al., 2006]), the discussion of issues pertaining to the selection of an optimal cortical parcellation is beyond the scope of this chapter.

Next, let model $\mathcal{M} = \{\lambda, \boldsymbol{\gamma}\}$ be the set of hyperparameters that encode the generative model proposed above. We use the Bayes theorem to express the posterior pdf of the sources given the data and model as,

$$p(\mathbf{x}_k|\mathbf{y}_k, \mathcal{M}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k, \mathcal{M})p(\mathbf{x}_k|\mathcal{M})}{p(\mathbf{y}_k|\mathcal{M})} \tag{2.8}$$

Note that $p(\mathbf{y}_k|\mathbf{x}_k, \mathcal{M}) \triangleq p(\mathbf{y}_k|\mathbf{x}_k, \lambda)$ and $p(\mathbf{x}_k|\mathcal{M}) \triangleq p(\mathbf{x}_k|\gamma)$ because the likelihood and priors are independent of $\boldsymbol{\gamma}$ and λ respectively. The density function $p(\mathbf{y}_k|\mathcal{M})$ is known as the model evidence [MacKay, 2008b] and, as we show in Section 2.2.2 (and in Chapter 3 in greater detail), its optimization allows us to reshape our modeling assumptions in a data-driven manner. The graph in Fig. 2.2 summarizes our generative model.

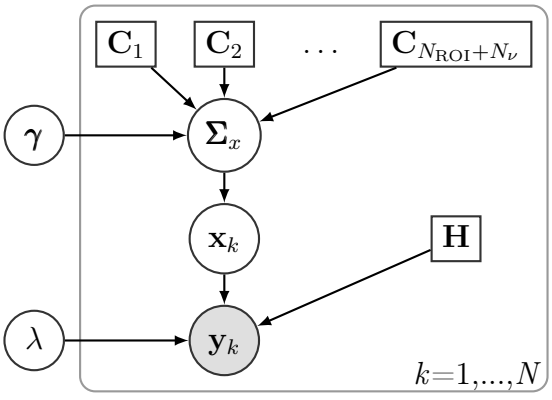


Figure 2.2. Graphical representation of our probabilistic generative model. Square, circle, and shaded circle symbols represent constant, hidden, and measured quantities respectively.

The model evidence is the normalization constant of the posterior in Eq (2.8),

therefore it can be ignored while searching for the maximum a posteriori (MAP) source estimate $\hat{\mathbf{x}}_k = \mathbf{x}_{MAP}$, which we find as the mode of the numerator of Eq (2.8) conditioned on $\mathcal{M}$,

$$\hat{\mathbf{x}}_k = \arg\max_{\mathbf{x}_k} p(\mathbf{y}_k|\mathbf{x}_k, \mathcal{M}) p(\mathbf{x}_k|\mathcal{M}) \tag{2.9}$$

We readily determine the functional form of the posterior as $p(\mathbf{x}_k|\mathbf{y}_k, \mathcal{M}) = N(\mathbf{x}_k|\hat{\mathbf{x}}_k, \boldsymbol{\Sigma}_{x|y})$, with the following conditional mean and covariance [Wipf and Nagarajan, 2009],

$$\begin{aligned} \hat{\mathbf{x}}_k &= \boldsymbol{\Sigma}_x \mathbf{H}^T \boldsymbol{\Sigma}_y^{-1} \mathbf{y}_k \\ \boldsymbol{\Sigma}_{x|y} &= \boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_x \mathbf{H}^T \boldsymbol{\Sigma}_y^{-1} \mathbf{H} \boldsymbol{\Sigma}_x \end{aligned} \tag{2.10}$$

where the model data covariance is given by the following expression,

$$\boldsymbol{\Sigma}_y = \lambda \mathbf{I}_{Ny} + \mathbf{H} \boldsymbol{\Sigma}_x \mathbf{H}^T \tag{2.11}$$

In case that we need to obtain the cleaned EEG signal $\bar{\mathbf{y}}_k$, e.g. for visualization or scalp ERP analysis, we subtract the artifact signal from the data as follows

$$\bar{\mathbf{y}}_k = \mathbf{y}_k - \mathbf{A}\hat{\boldsymbol{\nu}}_k = \mathbf{F}\mathbf{y}_k \tag{2.12}$$

where $\mathbf{F} = \mathbf{I} - \mathbf{A}\boldsymbol{\Sigma}_\nu \mathbf{A}^T \boldsymbol{\Sigma}_y^{-1}$ is a spatial filtering operator and $\hat{\boldsymbol{\nu}}_k$ is obtained (if needed) by selecting the last N_ν elements of the vector $\hat{\mathbf{x}}_k$. Likewise, the estimated PCD vector $\hat{\mathbf{g}}_k$ can be obtained by selecting the first N_g elements of $\hat{\mathbf{x}}_k$ or using the formula

$$\hat{\mathbf{g}}_k = \boldsymbol{\Sigma}_g \mathbf{L}^T \boldsymbol{\Sigma}_y^{-1} \mathbf{y}_k \tag{2.13}$$

and the source activity specific to the the ith ROI can be obtained using the formula

$$\hat{\mathbf{g}}_{i,k} = \gamma_i \mathbf{C}_i \mathbf{L}^T \boldsymbol{\Sigma}_y^{-1} \mathbf{y}_k \tag{2.14}$$

Further analysis of the source time series (e.g. ERP and connectivity analysis) can be done by averaging the source activity obtained in Eq (2.14) within ROIs,

$$\bar{g}_{i,k} = \frac{1}{N_i} \sum_{j \in \text{ROI}_i} \hat{\mathbf{g}}_{j,k} \tag{2.15}$$

where $\text{ROI}_i \subset \{1, \ldots, N_g\}$ is the subset of indices that belong to the ith ROI.

2.2.2 Model learning

The source estimates and cleaned data can be obtained analytically by evaluating the formulas given in Eq (2.10)-(2.13). These formulas, however, are model dependent because they depend on the specific values of the hyperparameters λ and $\boldsymbol{\gamma}$. In this section we outline the algorithm for learning those.

To evaluate Eq (2.10) conditioned on the optimal model estimate we maximize the posterior density,

$$\hat{\mathcal{M}} = \arg\max_{\lambda, \boldsymbol{\gamma}} p(\mathcal{M}|\mathbf{y}_k) \tag{2.16}$$

where $p(\mathcal{M}|\mathbf{y}_k) \propto p(\mathbf{y}_k|\mathcal{M})p(\mathcal{M})$ and $p(\mathcal{M})$ is a hyperprior. To determine the evidence described in Eq (2.8) we need to marginalize out the sources,

$$p(\mathbf{y}_k|\mathcal{M}) = \int p(\mathbf{y}_k|\mathbf{x}_k, \mathcal{M}) p(\mathbf{x}_k|\mathcal{M}) d\mathbf{x}_k \tag{2.17}$$

which for a linear Gaussian model like ours, is readily expressed as [Barber, 2012],

$$p(\mathbf{y}_k|\mathcal{M}) \propto \frac{\exp\left(-\frac{1}{2}\mathbf{y}_k^T \boldsymbol{\Sigma}_y^{-1} \mathbf{y}_k\right)}{|\boldsymbol{\Sigma}_y|^{-1/2}} \tag{2.18}$$

Next, we need to specify the hyperprior $p(\mathcal{M})$. Assuming that λ and γ_i are independent yields the factorization $p(\mathcal{M}) = p(\lambda) \prod_i p(\gamma_i)$. Since $\mathcal{M}$ contains only scale hyperparameters, a popular choice is to assume Gamma hyperpriors on a log-scale of λ^{-1} and γ_i^{-1}

[Tipping, 2001]. When the scale and shape parameters of the Gamma tend to zero the hyperpriors become flat (noninformative), in which case the optimization of the model posterior density depends only on the model evidence. This choice of hyperprior has the effect of assigning a high probability to low values of γ_i, which tends to shrink the irrelevant components of $\mathbf{x}_k$ to zero, leading to a sparsifying behavior know as Automatic Relevance Determination (ARD) [MacKay, 1992; Neal, 1996].

To obtain robust hyperparameter estimates in an online fashion, we use a block of consecutive measurements $\mathbf{Y} = [\mathbf{y}_k, \ldots, \mathbf{y}_{k+N}]$ as opposed to a single data point $\mathbf{y}_k$. Here we assume that the parameters characterizing the generative model vary at a slower rate compared to the brain and artifact dynamics producing the measurements. In other words, sources producing a given data block $\mathbf{Y}$ are not expected to go silent from one sample to the next. This assumption can be further motivated by the fact that the EEG can be temporally segmented into a sequence of discrete quasi-stable microstates, each of which consisting of a scalp configuration lasting for approximately 80 to 120 milliseconds before transitioning to a different microstate [Khanna et al., 2015; Koenig et al., 2002; Van De Ville et al., 2010]. With this in mind, we ignore short-term correlations within a $\mathbf{Y}$ block (iid assumption) and approximate the evidence of the ensemble as,

$$p(\mathbf{Y}|\mathcal{M}) \approx \prod_{k=1}^{N} p(\mathbf{y}_k|\mathcal{M}) \tag{2.19}$$

where N is set so that we deal with consecutive blocks of approximately 40 milliseconds in a way that we don't miss microstates.

The maximization of the model block evidence is equivalent to minimizing the so-called type-II Maximum Likelihood (ML-II) cost function [Barber, 2012], which is

obtained by applying $-2\log(\cdot)$ to Eq (2.19),

$$\mathcal{L}(\mathcal{M}) = \underbrace{\log|\boldsymbol{\Sigma}_y|}_{\text{Complexity}} + \underbrace{\text{trace}\left(\mathbf{C}_y\boldsymbol{\Sigma}_y^{-1}\right)}_{\text{Accuracy}} \tag{2.20}$$

where $\mathbf{C}_y = N^{-1}\mathbf{Y}\mathbf{Y}^T$ is the empirical data covariance. Note that the data should be zero-mean, which can be achieved with a high-pass filtering stage during preprocessing. Eq (3.6) embodies a tradeoff between model complexity and accuracy. Geometrically, the complexity term represents the volume of an ellipsoid defined by Σ_y. In particular, as the axes of the ellipsoid shrink due to the pruning of irrelevant sources, the volume is reduced. The second term measures model accuracy; i.e., how similar are the empirical and analytic covariances $\mathbf{C}_y$ and $\boldsymbol{\Sigma}_y$. We update the model on every block by solving the following optimization problem

$$\hat{\lambda}, \hat{\boldsymbol{\gamma}} = \arg\max_{\lambda,\boldsymbol{\gamma}} \mathcal{L}(\mathcal{M}) \tag{2.21}$$

We will show in Chapter 3 that Eq (2.21) can be solved efficiently using a two-stage optimization algorithm.

2.2.3 Connections between PEB+ and ICA

In the analysis presented above, the matrix $\mathbf{H}$ is prespecified. In this section, we analyze the generative model of Eq. (2.3) from the ICA viewpoint. ICA is a blind source separation method that seeks to estimate the source time series (often called activations in the ICA literature) $\mathbf{x}_k$ from the data time series $\mathbf{y}_k$ *without* knowing the gain (mixing) matrix $\mathbf{H}$. In ICA, we assume that the latent sources are instantaneously independent, which yields the following prior distribution

$$p(\mathbf{x}_k) = \prod_{i=1}^{N_x} p_i(x_{i,k}) \tag{2.22}$$

To simplify the exposition, we assume the same number of sensors and sources,

$N_y = N_x$, and the interested reader can find the case $N_y < N_x$ in Le et al. [2011]; Lewicki and Sejnowski [1998]. From these premises, the objective of the algorithm is to learn the unmixing matrix $\mathbf{H}^{-1}$ such that we can estimate the sources with $\hat{\mathbf{x}}_k = \hat{\mathbf{H}}^{-1}\mathbf{y}_k$. The unmixing matrix $\hat{\mathbf{H}}^{-1}$ can be learned up to a permutation and rescaling factor, which has the inconvenience that the order of the learned components can change depending on the starting point of the algorithm and data quality. We can use a data block $\mathbf{Y}$ to write the likelihood function

$$p(\mathbf{Y}|\mathbf{H}, \lambda) = \prod_{k=1} p(\mathbf{y}_k|\mathbf{H}, \lambda) \tag{2.23}$$

under the assumption of independent data collection. However, we should point out that in ICA, $\mathbf{Y}$ is usually a data block longer than the one considered in Chapter 3, thus the iid data assumption is harder to justify. To alleviate this situation the data are usually whitened during preprocessing. We can obtain each factor in Eq (2.23) by integrating out the sources as follows

$$p(\mathbf{y}_k|\mathbf{H}, \lambda) = \int p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{H}, \lambda)p(\mathbf{x}_k)d\mathbf{x}_k \tag{2.24}$$

As noted by MacKay [2008a], assuming that the data are collected in the noiseless limit, $\lambda \to 0$, transforms the Gaussian likelihood $p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{H}, \lambda)$ into a Dirac delta function, in which case Eq (2.24) leads to the Infomax algorithm of Bell and Sejnowski [1995]. The learning algorithm essentially consists in finding the gradient of the log likelihood, $\log p(\mathbf{Y}|\mathbf{H}, \lambda)$, with respect to $\mathbf{H}$ and updating $\mathbf{H}$ on every iteration such that the probability of the data increases. As pointed out by Comon [1994], the ICA model is uniquely identifiable only if at most one component of $\mathbf{x_k}$ is Gaussian. Therefore, the prior densities $p_i(x_{i,k})$ are usually assumed to exhibit heavier tails than the Gaussian and, in particular, the prior $p_i(x_{i,k}) \propto \cosh^{-1} x_{i,k}$ yields the popular ICA contrast function $\tanh(\mathbf{H}^{-1}\mathbf{y}_k)$. Note that this prior is not motivated by a biological consideration but by a mathematical

necessity.

It is remarkable that ICA can learn columns of $\hat{\mathbf{H}}$ that are consistent with bipolar (single or bilaterally symmetric) cortical current source scalp projections without using any anatomical or biophysical constraint whatsoever [Makeig et al., 1997]. Onton et al. [2006] have shown that other columns may correspond to different stereotypical artifact scalp projections as well as a set of residual scalp maps that are difficult to explain from a biological standpoint. Delorme et al. [2012] have shown that the best ICA algorithms can identify approximately 30% of dipolar brain components (approximately 21 brain components out of 71 possible in a 71-channel montage). Although ICA has proven to be a useful technique for the study of brain dynamics [Makeig and Onton, 2011], we must wonder if its performance can be improved, perhaps by making BSS of EEG data less "blind". In other words, if we know a priori what kind of source activity we are looking for (dipolar cortical activity, EOG and EMG artifacts and so on), why limit ourselves to a purely blind decomposition?

In this chapter, we advocate the use of as much information as we can to help solve the ill-posed inverse problem. In that sense, the use of a prespecified lead field matrix in the generative model of the EEG forces inverse algorithms to explain the data in terms of dipolar sources, because the lead field is precisely an overcomplete dictionary of dipolar projections of every possible source there is in a discretized model of the cortex. It has been shown that source estimation can greatly benefit from the use of geometrically realistic subject-specific [Cuspineda et al., 2009] or, alternatively, population-based approximated lead fields matrices [Valdés-Hernández et al., 2009]. Furthermore, augmenting the lead field dictionary with a set of stereotypical artifact projections, as proposed in Section 2.2.1, furnishes a more realistic generative model of the EEG in a way that renders blind decomposition unnecessary or at least suboptimal for brain imaging.

2.2.4 Independent components through PEB+

Since the source activity measured in the EEG is mixed by the volume conduction effect, ideally, we would like the PEB+ framework to exhibit the ICA property of yielding maximally independent (demixed) source time series. In this section we show that this is indeed the case. We start by rewriting the biologically motivated source prior of Eq (2.5) as

$$p(\mathbf{x}_k|\gamma) = \prod_{i=1}^{N_\gamma} p_i(\mathbf{x}_{i,k}|\gamma_i) \tag{2.25}$$

where each factor is a Gaussian pdf and i indexes a group of sources or an artifact component. To write Eq (2.25) as the ICA prior in Eq (2.22) we need to integrate out the hyperparameter γ_i from each factor as follows:

$$p(\mathbf{x}_k) = \prod_{i=1}^{N_\gamma} \underbrace{\int p_i(\mathbf{x}_{i,k}|\gamma_i)p(\gamma_i)d\gamma_i}_{p_i(\mathbf{x}_{i,k}) \text{ is a Student t-distribution}} \tag{2.26}$$

which, given our choice of hyperprior on γ_i, renders each marginalized prior $p_i(\mathbf{x}_{i,k})$ a heavy-tailed Student t-distribution [Tipping, 2001]. We note that in our development we take the route of optimizing the γ_i hyperparameters rather than integrating them out because the former approach yields a simpler algorithm and tends to produce more accurate results in ill-posed inverse problems [MacKay, 1996]. Moreover, the optimization of γ_i allows for automatic removal of irrelevant brain and artifact components that are not supported by the data, thereby eliminating the subjectivity implicit in manual component selection. Assuming the prior in Eq (2.25), the ICA data likelihood of Eq (2.23) becomes exactly the evidence of Eq (2.19), with the difference that in the PEB+ algorithm the $\mathbf{H}$ matrix is known and the evidence is optimized on small blocks of data, which gives our algorithm the ability to run in an online manner and to capture transient brain dynamics.

We summarize the advantages of using the PEB+ framework over ICA for source

separation and imaging of EEG data as follows:

- It deals gracefully with the overcomplete case ($N_y \ll N_x$) by finding the MAP source estimator, which always exist even in the presence of rank-deficient data, e.g. after removing the common average reference.
- It deals with the redundancy in brain responses by inducing independence over groups of sources.
- The use of the ARD prior allows for the automatic selection of components in a data-driven manner, thereby eliminating the subjectivity of selecting components based on practitioner's experience.
- It can adapt to non-stationary dynamics by updating the model on smaller blocks of data.
- It can be used in online applications by leveraging fast evidence optimization algorithms.
- Artifact removal, source separation, and imaging can be obtained simultaneously as a consequence of optimizing the evidence of a biologically informed generative model.
- It facilitates subject-level analysis because we estimate the same number of cortical source activations per subject, each of which has known anatomical support. This eliminates the complications of clustering ICs and dealing with missing components [Bigdely-Shamlo et al., 2013b] while allowing the use of more straightforward and widespread statistical parametric mapping techniques [Penny et al., 2007].

2.3 Results

The PEB framework has been validated extensively on simulated and real data elsewhere [Friston et al., 2008; Henson et al., 2011; Ojeda et al., 2018; Zhang and Rao, 2013]. In this section, we study the effects of modeling artifactual sources in the quality of the model inversion. We also discuss source separation through PEB+, as well as its use for data cleaning. Then in Section 2.3.5 we show an application of PEB+ to the study of heading computation during full-body rotations in the context of a mobile brain/body imaging (MoBI) experiment.

2.3.1 Construction of artifact projection dictionary A

To characterize artifactual ICs we used data from two different studies made public under the umbrella of the BNCI Horizon 2020 project[1] [Brunner et al., 2015]. Since in the next two sections we investigate different features of the PEB+ algorithm rather than the biological interpretation of its results, we don't dwell into the details of the experimental paradigms used in each study, and we direct the interested reader to the respective publications referenced below.

Data set 1: Error related potentials

The first study, *013-2015*, provided EEG data from 6 subjects (2 independent sessions per subject and 10 blocks per session) collected by Chavarriaga and del R. Millán [2010] using an experimental protocol designed to study error potentials during a BCI task. EEG samples were acquired at a rate of 512 Hz using a Biosemi ActiveTwo system and a 64-channels montage placed according to the extended 10/20 system.

[1] http://bnci-horizon-2020.eu/database/data-sets

Data set 2: Covert shifts of attention

The second data set, *005-2015*, provided EEG and EOG data from 8 subjects collected by Treder et al. [2011] using an experimental protocol designed to study the EEG correlates of shifts in attention. The EEG was recorded using a Brain Products actiCAP system, digitized at a sampling rate of 1000 Hz. The montage employed had 64 channels placed according to the 10/10 system referenced to the nose. In addition, an EOG channel (labeled as EOGvu) was placed below the right eye. To measure vertical and horizontal eye movements, from the total of 64 EEG channels, two were converted into bipolar EOG channels by referencing Fp2 against EOGvu, and F10 against F9, thereby yielding a final montage of 62 EEG channels.

Data preprocessing and IC scalp maps clustering

After transforming each data file to the *.set* format, both studies were processed using the same pipeline written in MATLAB (R2017b The MathWorks, Inc., USA) using the EEGLAB toolbox [Delorme et al., 2011]. The pipeline consisted of a 0.5 Hz high-pass forward-backward FIR filter and re-referencing to the common average, followed by the Infomax ICA decomposition of the continuous data. We pooled all the preprocessed data sets and randomly assigned them to one of two groups: 80 % to the training set and 20 % to the test set. The training set was used to construct the artifact dictionary and the test set was used to evaluate the performance of the PEB+ algorithm.

To construct an artifact dictionary from a heterogeneous EEG database, we need to represent each independent scalp map into a common (co-registered) channel space. To that end, we used the coordinates of the common channels between the montages used in the two studies described above to estimate a linear transformation from the 62-channel space to the 64-channel one. After co-registration, we pooled both studies in the training set resulting in a matrix of 64 channels by 6774 independent scalp maps (101 sessions and blocks yielding 64 ICs each plus 5 sessions yielding 62 ICs each). It is worth emphasizing

that we only warped IC scalp maps and not the actual data or IC activations.

Next, we used the matrix of co-registered independent scalp maps to estimate clusters using the k-means algorithm. Clusters were labeled as Brain, EOG, EMG, or Unknown (scalp maps of unknown origin) by an expert. Unknown clusters were not used further in this chapter. Fig 2.3 shows a visualization of the IC scalp maps using the t-distributed stochastic neighbor embedding (t-sne) algorithm [Van Der Maaten and Hinton, 2008]. The t-sne algorithm allows us to represent each 64-dimensional IC scalp map as a dot in a 2D space in a way that similar and dissimilar scalp maps are modeled by nearby and distant points respectively with high probability. We ran the k-means algorithm for several numbers of clusters, and we stopped at 13 after noticing that many small islands scattered at the periphery of Fig 2.3 started to be either mislabeled as Brain or labeled consistently as EOG, EMG or Unknown. The grey points in the figure denote most of the scalp maps labeled as non-brain.

Using the insights from Fig. 2.3, we completed the augmented PGM of Eq (2.3) by building the $\mathbf{A}$ dictionary as follows:

$$\mathbf{A} = \left[\mathbf{a}_{EOG_v}, \mathbf{a}_{EOG_h}, \mathbf{a}_{EMG_1}, \ldots, \mathbf{a}_{EMG_{11}}, \mathbf{I}_{N_y}\right] \tag{2.27}$$

where $\mathbf{a}_{EOG_v}$ and $\mathbf{a}_{EOG_h}$ are the centroids of the vertical and horizontal EOG clusters respectively, $\mathbf{a}_{EMG_i}$ are the centroids of EMG clusters and we modeled spike artifacts affecting each individual channel with the columns of the identity matrix $\mathbf{I}_{N_y}$, with $N_y = 64$.

Calculation of subject-specific H matrices

Depending on the montage of each subject, we nonlinearly warped their 62 or 64-channel montage to the scalp surface of a four-layer (scalp, outer skull, inner skull, and cortex) "Collin27" template using the DSI toolbox. Then we computed the orientation-fixed lead field matrices $\mathbf{L}$ using the boundary element method solver in the OpenMEEG toolbox

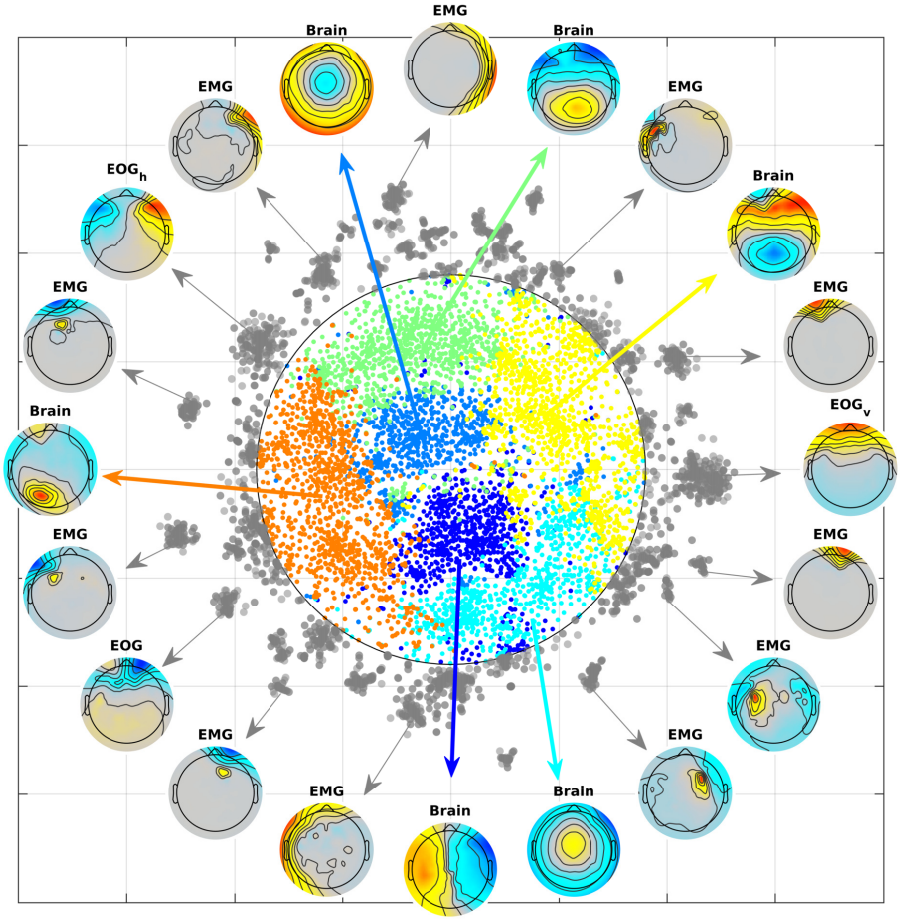


Figure 2.3. t-sne visualization of IC scalp map clusters. We used the t-sne algorithm to represent each 64-dimensional scalp map as a dot in a 2D space in a way that similar and dissimilar scalp maps are modeled by nearby and distant points respectively with high probability. The clusters were estimated using the k-means algorithm. The grey points indicate mostly non-brain or mislabeled scalp projections.

[Gramfort et al., 2010]. Next, we calculated individual $\mathbf{A}$ matrices by linearly warping its columns from the 64-channel space to the space defined by the head surface of the template. Finally, we divided each column of the augmented dictionary $\mathbf{H}$ by its norm so that their relative contribution to the scalp EEG could be determined by the amplitude of the source activation vector $\mathbf{x}_k$.

2.3.2 PEB+ artifact model validation

In this section, we investigate whether the approach of explicitly modeling artifact scalp projections used in PEB+ yields significantly better source estimates than the traditional PEB. To quantify the support in the data for a given model we used the Bayes factor. In Bayesian model selection/comparison, the Bayes factor is used as an alternative to classical hypothesis testing, replacing the p-value as a measure of evidential strength while avoiding the abuse to which the latter is often subjected to these days [Stern, 2016]. The Bayes factor between two generative models $\mathcal{M}_i$ and $\mathcal{M}_j$ is denoted as $B_{i,j}$ and is defined as the ratio between the evidence of each model:

$$B_{i,j} = \frac{p(\mathbf{Y}|\mathcal{M}_i)}{p(\mathbf{Y}|\mathcal{M}_j)} \tag{2.28}$$

Interpreting $B_{i,j}$ is straightforward, for instance $B_{i,j} > 1$ indicates that there is more evidence in favor of generative model $\mathcal{M}_i$ over $\mathcal{M}_j$. Likewise, $B_{i,j} \approx 1$ means that there is no conclusive evidence in favor of any of the models considered. Usually, a Bayes factor in $2\log_e$ units higher than 2, 6, and 10 is respectively considered positive, strong and very strong evidence in favor of $\mathcal{M}_i$ [Kass and Raftery, 1995]. Conversely, note that Eq (2.28) can be interpreted as evidence in favor of $\mathcal{M}_j$ simply by flipping the ratio.

As we explained earlier, in this chapter we characterize artifacts empirically rather than in a mechanistic principled way. Therefore, next we used Bayes factors to assess the performance of the PEB+ algorithm under different variants of the artifact dictionary $\mathbf{A}$. In our analysis we considered the following models:

- $\mathcal{M}_0$: ignoring artifacts, $\mathbf{A} \in \emptyset$ (classic PEB)
- $\mathcal{M}_1$: modeling artifacts with Eq (2.27),
- $\mathcal{M}_2$: modeling EOG and EMG components only, $\mathbf{A} = [\mathbf{a}_{EOG_v}, \mathbf{a}_{EOG_h}\mathbf{a}_{EMG_1}, \ldots, \mathbf{a}_{EMG_{11}}]$,

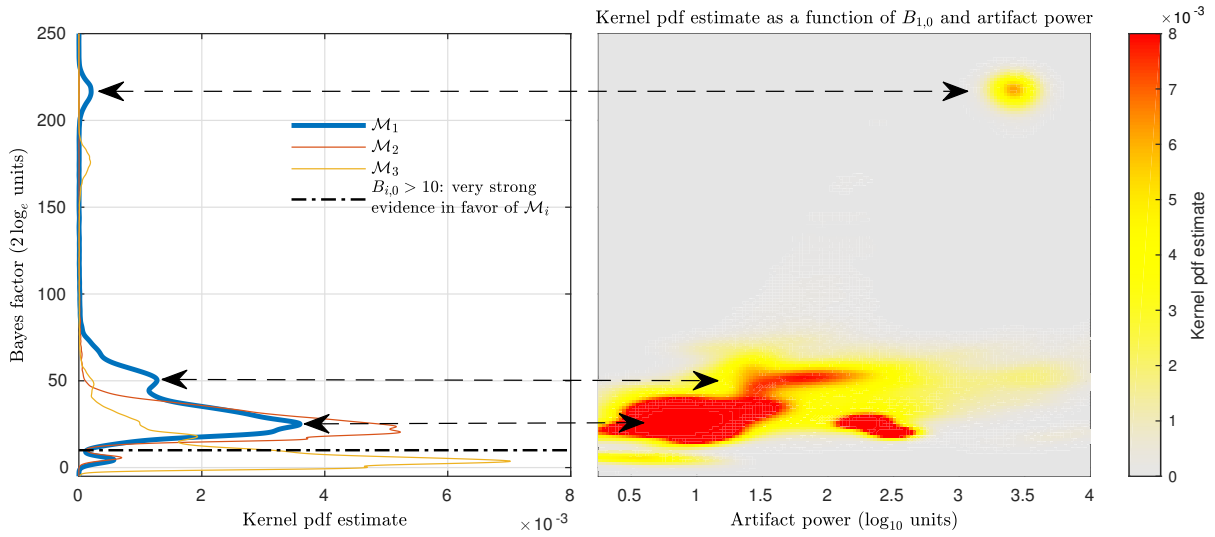


Figure 2.4. Comparison of generative models with different variants of the artifact dictionary with respect to $\mathcal{M}_0$. **Left:** Bayes factor kernel pdf estimate for each artifact model. **Right:** $B_{1,0}$ kernel pdf estimate as a function of artifact power.

- $\mathcal{M}_3$: modeling single channel spikes only, $\mathbf{A} = \mathbf{I}_{N_y}$.

To that end, we estimated the brain and artifact source time series of all subjects in the testing set under each model. We collected the log evidence afforded by each model in blocks of 40 msec of data and computed Bayes factors with respect to $\mathcal{M}_0$.

The left panel of Fig 2.4 shows the kernel pdf estimate of each Bayes factor: $B_{1,0}$, $B_{2,0}$, $B_{3,0}$ in blue, orange, and yellow traces respectively. The area above the dot-dashed black trace represents the probability space in which modeling artifacts yielded a better generative model for the EEG signal. We computed the probability of having very strong evidence in favor of model $\mathcal{M}_i$ by integrating over Bayes factors higher than 10, $P_i(10 \leq B_{i,0}) = \int_{B_{i,0} \geq 10} f_i(B_{i,0})$, where $f_i(B_{i,0})$ denotes a pdf as a function of model $\mathcal{M}_i$. The models 1 to 3 scored 0.9704, 0.9705, and 0.4202. Since models $\mathcal{M}_1$ and $\mathcal{M}_2$ yielded the highest probability and because $\mathcal{M}_1$ includes $\mathcal{M}_2$ and $\mathcal{M}_3$, henceforth we use the artifact dictionary given by Eq (2.27).

To further illustrate the importance of modeling artifact components, the right panel of Fig 2.4 shows the kernel pdf estimate of $B_{1,0}$ as a function of the artifact power.

The artifact power was calculated as the maximum RMS power over the artifact sources for each 40 msec window. We note that the shape of its pdf depicted on the left panel seems to be determined by the higher performance of the PEB+ algorithm under different clusters of artifactual activity expanding several orders of magnitude.

Fig 2.5 shows an example of applying PEB+ to an epoch of data of 4 seconds centered around an eye blink event. Panel **A** shows a 32-channel subset of the raw and reconstructed (cleaned) EEG traces in black and red respectively. In addition to an eye blink, we have a lateral eye movement event around the -1 sec latency. Panel **B** shows the estimated EOG_v and EOG_h artifact source activity in blue and orange respectively. We note that these artifact sources are active only at the latencies where the EEG is affected and mostly zero elsewhere. In panel **C**, the first and last two columns represent the raw and cleaned EEG topographies at the maximum of the lateral eye movement and eye blink events respectively. Panel **D** shows different views of the estimated cortical source maps underlying the raw topographies in **C**. The cleaned topographies in **C** are obtained after the estimated artifact sources are projected out of the data. Panel **E** shows the log evidence for generative models $\mathcal{M}_1$ and $\mathcal{M}_0$ in blue and orange respectively. Note that both traces differ mostly only when artifacts occur and higher log evidence in favor of model $\mathcal{M}_1$ indicates that source estimation benefits from modeling artifacts.

It is worth noting that, in the last column of panel **D**, some residual eye blink artifact seems to be mistakenly represented as a small activation in the frontal pole. We point out that, in practice, it may be extremely hard to totally remove artifactual activity because: 1) the use of a lead field matrix derived from a template head model may misfit the anatomy of the subject introducing errors in the **L** dictionary, 2) errors in the sensor locations can cause the EEG topography to shift with respect to the expected brain and artifact source projections, 3) EMG scalp projections are difficult to characterize due to their variability, as opposed to EOG projections that are more stereotyped, and 4) unmodeled muscle projections, such as those towards the back of the head that were largely

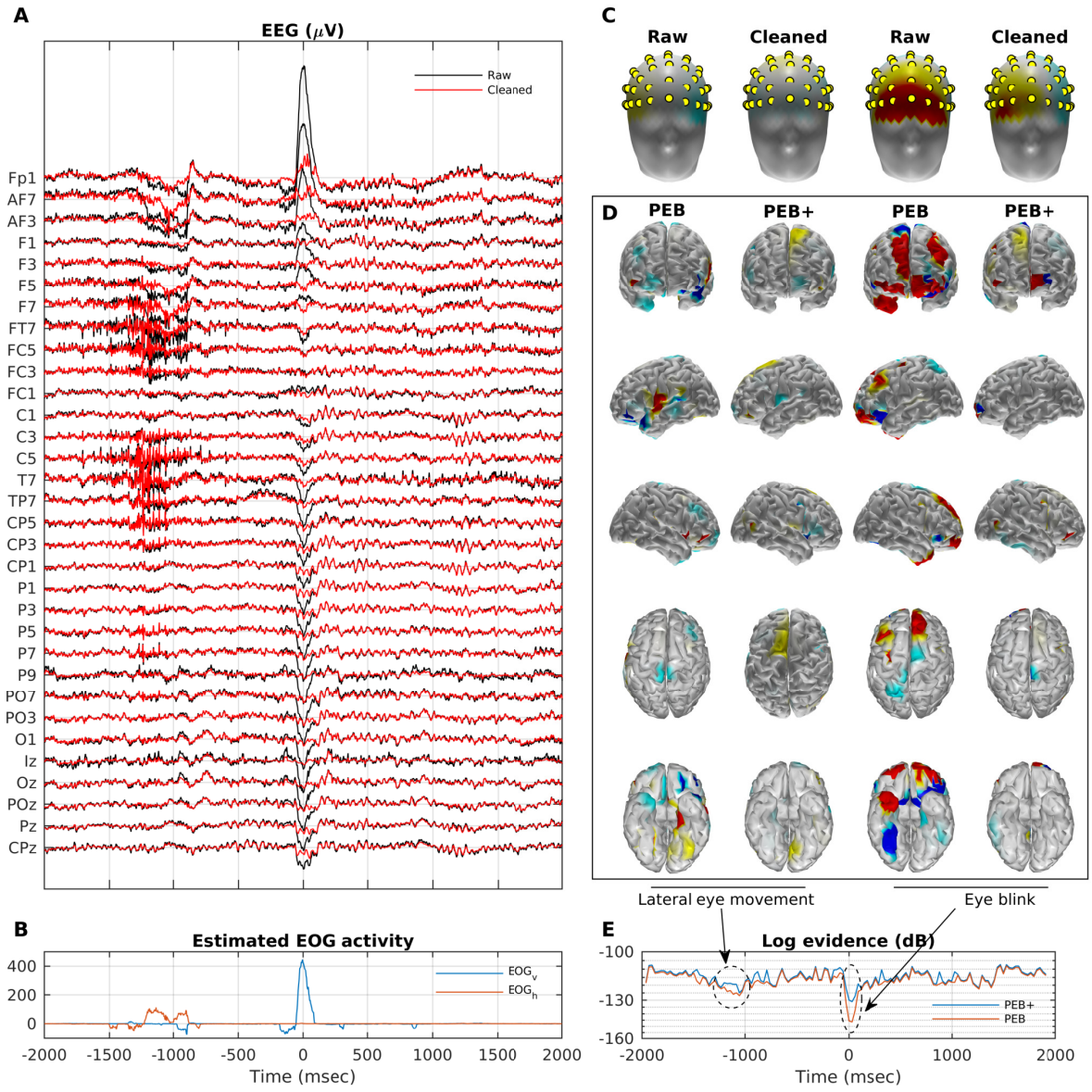


Figure 2.5. Example of applying PEB+ to an epoch of EEG with lateral eye movement and eye blink artifacts. **A**: 32-channel subset of raw and cleaned EEG traces. **B**: Estimated EOG_v and EOG_h artifact source activity. **C**: Columns 1 and 3 and 2 and 4 represent the raw and cleaned EEG topographies at the maximum of the lateral eye movement and eye blink events respectively. **D**: Different views of the estimated cortical source maps underlying the raw topographies in **C**. **E**: Log evidence yielded by PEB+ and PEB algorithms on consecutive 40 msec blocks of data along this epoch.

ignored in this study. Despite all these issues, Figures 2.4 and 2.5 demonstrate that PEB+ can yield reasonably robust source estimates in the presence of artifacts. Furthermore,

panel **E** of Fig 2.5 suggests that we could use dips in the log evidence to inform subsequent processing stages of artifactual events that were not successfully dealt with.

2.3.3 Source separation performance

In this section, we investigate the source separation performance of the PEB+ algorithm. To that end, we used the test set to compare PEB+ and Infomax ICA regarding 1) volume conduction unmixing performance and 2) data size requirements for good source separation. We assessed the unmixing performance by calculating the mutual information reduction (MIR) achieved by each algorithm on data blocks of different sizes.

The MIR is an information theoretic metric that measures the total reduction in information shared between the components of two sets of multivariate time series. The mutual information (MI) between two given time series $x_{i,k}$ and $x_{j,k}$, $I(x_i, x_j)$, can be defined as the Kullback-Leibler (KL) divergence between their joint and marginal distributions:

$$I(x_i, x_j) = D_{KL}[p(x_i, x_j) \parallel p(x_i), p(x_j)] \tag{2.29}$$

where $I(x_i, x_j) > 0$ indicates that processes x_i and x_j share information while $I(x_i, x_j) = 0$ indicates that they are statistically independent such that

$$p(x_i, x_j) = p(x_i)p(x_j)\cancelto{0}{p(x_i|x_j)}$$

We define the MIR of source separation algorithm A with respect to B, as the difference in normalized total pairwise MI (PMI) achieved by each decomposition:

$$\begin{aligned} MIR_{A,B} = & \frac{2}{N_A(N_A-1)} \sum_{i=2}^{N_A} \sum_{j=1}^{i-1} I(x_i^A, x_j^A) - \\ & \frac{2}{N_B(N_B-1)} \sum_{i=2}^{N_B} \sum_{j=1}^{i-1} I(x_i^B, x_j^B) \end{aligned} \tag{2.30}$$

where x_i^A and x_i^B are the set of components yielded by each method and N_A and N_B are the number of components afforded by each decomposition. We note that to obtain a PMI that is not biased by the number of components, we normalize each summation by the number of unique (i, j) pairs. Here we calculated the MI using the non-parametric kernel pdf estimates of the quantities in Eq (2.29) for the the multichannel EEG data, the ROI-collapsed sources estimated by PEB+, and the ICs obtained by Infomax.

In Fig 2.6, the left panel shows a box plot of the MIR of PEB+ and Infomax calculated with respect to the MI of channel data. As indicated by the x-axis, we ran the experiment multiple times varying the data sizes from 0.5 to 500 seconds ($\sim$ 8 minutes). As expected, both algorithms reduce source MI, thereby reversing to some extent the mixing effect of the volume conduction. We see also that, on average, when the MIR is calculated in short blocks of data, PEB+ exhibits higher unmixing performance while Infomax seems to do better on longer blocks. This effect is more clearly represented in the panel on the right, which shows the box plot of the MIR of PEB+ with respect to Infomax. In that panel, distributions with entire positive (orange) or negative (blue) values indicate a significant source crosstalk reduction performance in favor of the PEB+ or Infomax algorithms respectively. We note that PEB+ better captures transient dynamics for short 0.5-8 sec data blocks.

It is worth noting that with PEB+, it is possible to update the unmixing matrix (given by the term $\mathbf{\Sigma_x H^T \Sigma_y^{-1}}$ in Eq (2.10)) on a time-scale of tens of milliseconds because of the regularization induced by the multiple constraints. This allows for adaptation to non-stationary brain and artifact source dynamics. Infomax (and most ICA algorithms) on the other hand, requires larger data blocks to learn a global factorization of mixing matrix and source activations of reasonable quality. Moreover, Fig 2.6 suggests that the estimation of a global ICA model is suitable for identifying components that remain stationary over the whole experiment, but otherwise, it is suboptimal for capturing transient dynamics.

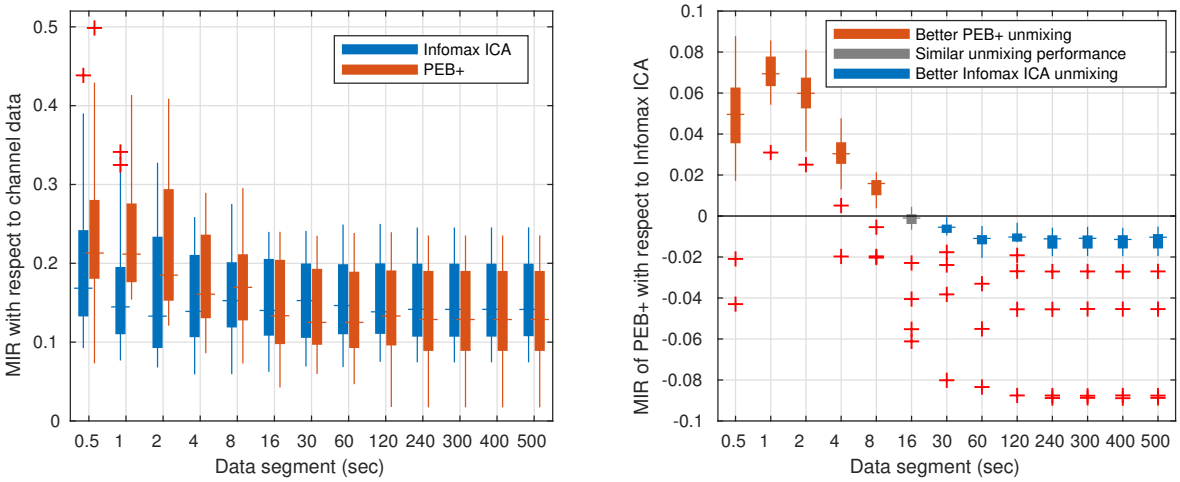


Figure 2.6. Source separation performance. **Left:** Box plot of MIR with respect to channel data computed on blocks of various sizes. **Right:** Box plot of MIR of PEB+ with respect to Infomax ICA for the same data blocks shown on the left. On each box, the central mark indicates the median, and the bottom and top edges indicate the 25th and 75th percentiles respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the + symbol. On the right, the distributions with entire positive (orange) or negative (blue) values indicate a significant source crosstalk reduction in favor of the PEB+ or Infomax algorithms respectively.

2.3.4 Data cleaning performance

In this section, we benchmark the data cleaning performance of the PEB+ algorithm against ASR. The ASR algorithm has gained popularity in recent years for its ability to remove a variety of high amplitude artifacts in an unsupervised manner, thereby enabling automatic artifact rejection for offline as well as real-time EEG-based BCI applications. Since in real data we do not have a ground truth for artifactual activity, we benchmark the methods according to the correlation between raw and cleaned data samples in blocks with negligible or no artifactual activity, where low correlation values indicate needless distortion of the brain activity.

We ran both algorithms for each subject in the test set and collected the following quantities on subsequent blocks of 40 msec: 1) the correlation between raw and cleaned data (computed as the correlation between the correspondent data blocks vectorized across channels and time points) and 2) the maximum RMS artifact power yielded by PEB+ as described in Section 2.3.2. ASR's performance depends on multiple parameters, but it has been indicated that the most critical one is the cutoff [Chang et al., 2018]. In the first experiment we used a cutoff equal to 5, which was the default value of EEGLAB's ASR plugin at the time of this publication.

In Fig 2.7, the left and right panels show the empirical kernel pdf estimation of the correlation as a function of the artifact's power for the ASR and PEB+ algorithms respectively. We see that in both methods, the correlation decreases as artifact power increases. This effect is expected and desired because cleaning algorithms are supposed to modify contaminated raw data. Towards low power artifact regions, however, ASR exhibits a significant amount of probability mass that spreads down to low correlation values while PEB+ seems to have most of its probability mass bounded from below at around 0.8. This result indicates that, at a cutoff of 5, ASR cleaning is overly aggressive to the point of significantly modifying the data in the absence of artifacts. These findings

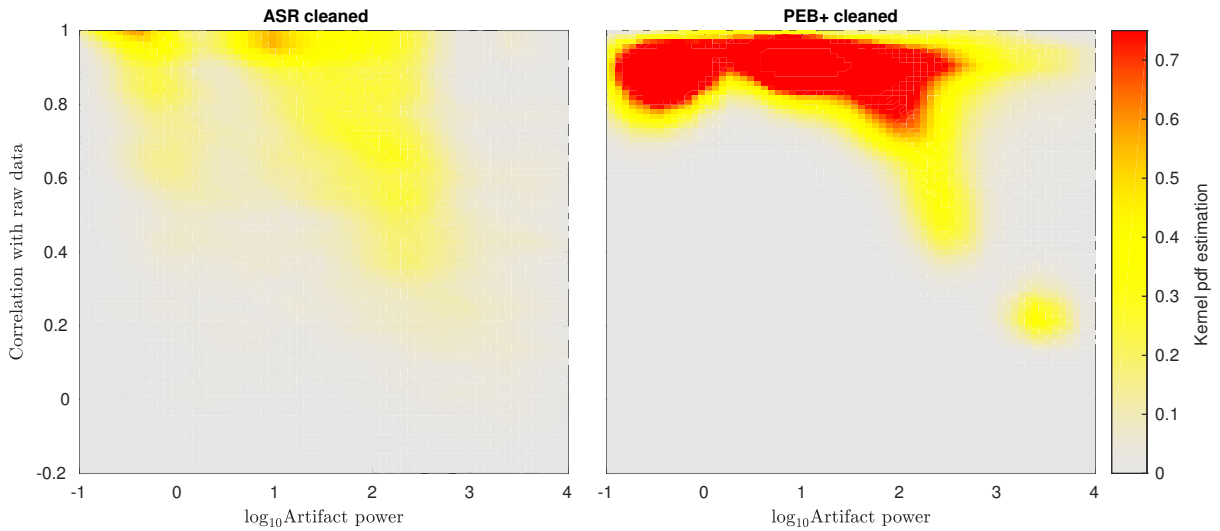


Figure 2.7. Data cleaning performance. Kernel pdf estimation of the correlation between raw and cleaned data as a function of artifact power. **Left:** Data cleaned by ASR using cutoff=5 (default). **Right:** Data cleaned by PEB+. Note that, as expected, in both algorithms the correlation drops as artifacts increase. Towards low amplitude artifacts, however, ASR significantly distorts the data while PEB+ does not.

are in agreement to what was recently reported by Chang et al. [2018].

In Section 2.3.2 we showed that PEB+ can reduce the effects of eye-related artifacts. In Fig 2.8 we show an example of its performance removing EMG. The data correspond to an excerpt of EEG extracted from a subject selected at random from the test set while he/she was performing the respective cognitive task. In the left panel, the gray and colored traces represent contaminated and cleaned EEG signals respectively. The traces shown correspond to channels located on each side of the cap. Channels in these areas are often contaminated by EMG activity due to their proximity to the temporalis muscle [Fu et al., 2006]. As we can see, the higher amplitude decorrelated EMG activity is largely reduced. In the panel on the right, the gray and colored traces represent the power spectral density estimates of their respective channels on the left. We see that a significant amount of broad band power related to the EMG activity was removed, especially towards frequencies higher than 18 Hz. We note that a strong 50 Hz AC line noise remains in the cleaned data, this is expected because in our approach we do not model this type of artifacts and

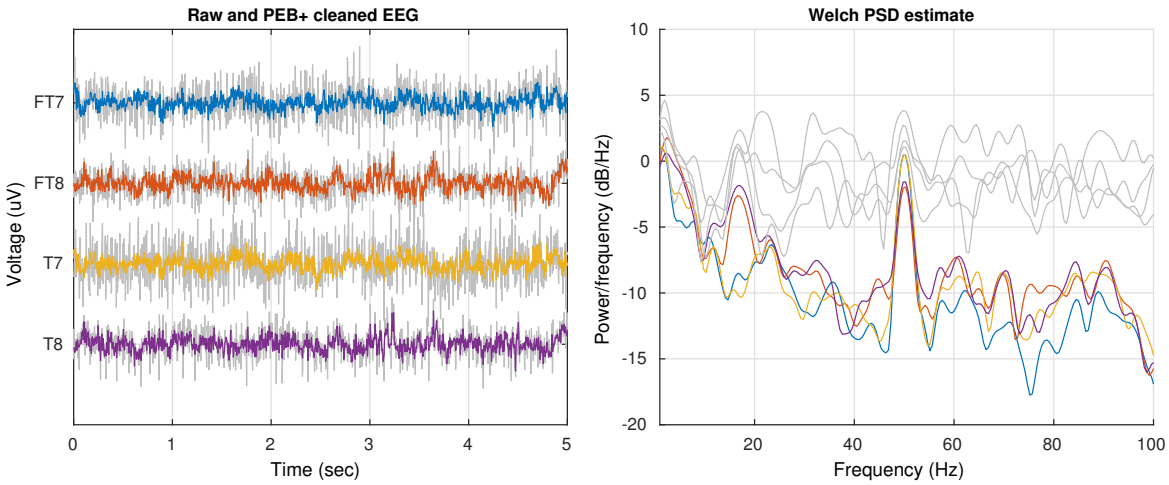


Figure 2.8. EMG artifact cleaning. EEG channels contaminated by EMG noise, the gray and colored traces represent raw and cleaned data respectively. **Left:** Excerpt of EEG signal. **Right:** Welch power spectral density estimation of the data shown on the left.

they can be relatively easily removed using a notch filter.

To further illustrate the results of the analysis shown in Fig 2.7, in Fig 2.9 we compare the cleaning performance of PEB+ against ASR using two cutoff parameters (5 and 100) and ICA on a typical epoch of EEG containing lateral eye movement, eye blink, and muscle artifacts. As before, the EEG data were extracted from a subject selected at random from the test set. In panel **A**, the shaded area in blue indicates a segment of clean data, while green, red, and yellow areas indicate segments contaminated by lateral eye movement, muscle, and eye blink artifacts respectively. The correlation achieved by each method (computed by vectorizing all channels and samples in the blue segment) is displayed on the top left corner of each panel. All correlations were significant with p-values lower than 0.005. Panels **A**, **B**, **C**, and **D** show the raw and cleaned EEG traces produced by PEB+, ICA, ASR (100), and ASR (5) methods respectively. For the ICA approach we cleaned the data by removing the contribution of several stereotypical EOG and EMG components selected manually. PEB+ and ICA displayed similar performance in the sense that lateral eye movement and eye blink artifacts were largely removed, EMG was not totally removed by PEB+, while the clean data segment was minimally distorted,

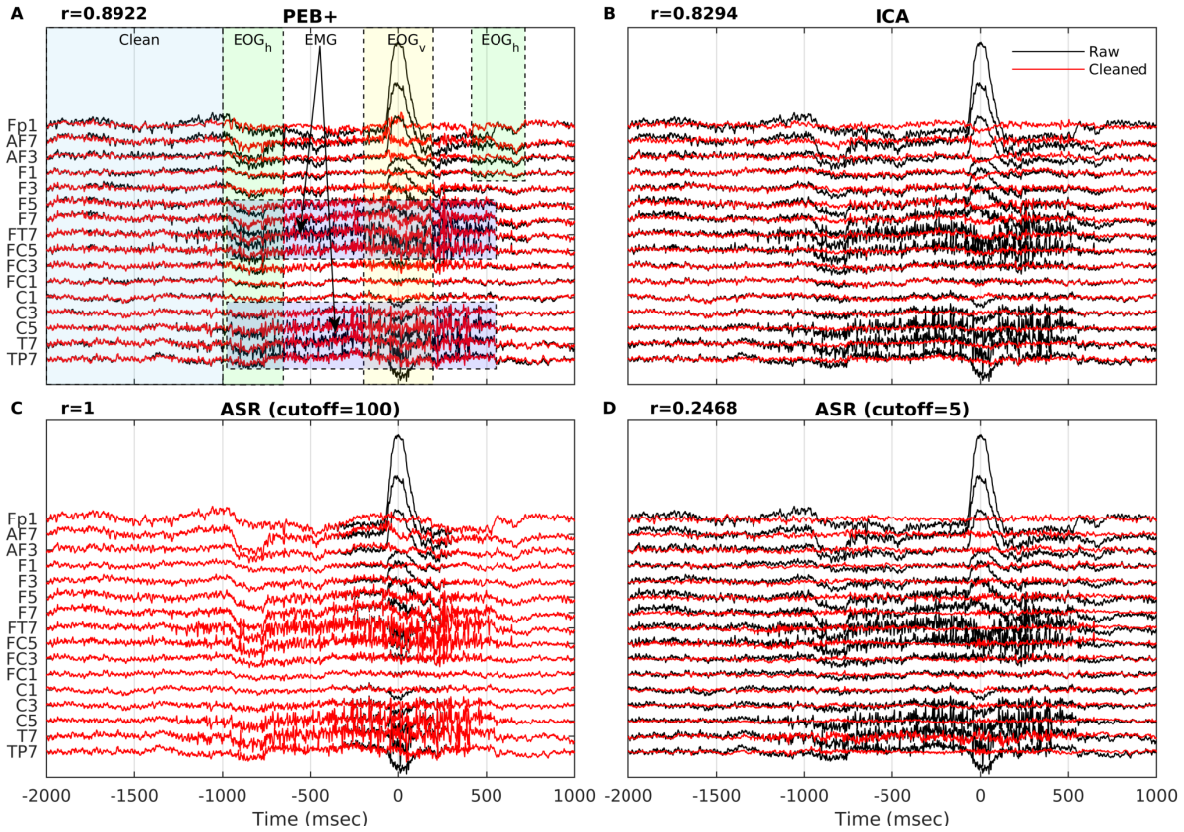


Figure 2.9. Example of the data cleaning performance of PEB+, ICA, and ASR on a noisy epoch. Only 16 channels are shown. **A**: PEB+. **B**: ICA. **C**: ASR with cutoff=100. **D**: ASR with cutoff=5 (current default value in EEGLAB's *clean_rawdata* plugin). In panel **A**, the segment of data used to compute the correlations between raw and cleaned EEG is indicated with the blue shaded area (-2000 msec to -1000 msec) while the areas shaded in green, yellow and red indicate lateral eye movement, eye blink, and muscle artifacts respectively.

as indicated by correlations with the raw samples of 0.8922 and 0.8294 respectively. We note that the distortion introduced by ICA could be reduced by a more conservative selection of the artifactual ICs to remove. ASR (100) did not distort the clean data segment and removed the higher amplitude eye blink artifact but failed to remove lateral eye movement and muscle artifacts. ASR (5) removed all the artifactual activity, however it also significantly distorted the clean data segment as indicated by a correlation of 0.2468.

2.3.5 Heading computation during full-body rotations

We finalize the chapter with an application of the PEB+ algorithm to MoBI data. MoBI experiments are notoriously difficult to analyze due to the amount of motion-induced artifacts as well as the presence of transient and stationary brain dynamics of variable duration across trials. Here, we try to replicate the main findings of a study that looked into the dynamics of the retrosplenial cortex (RSC) supporting heading computation during full-body rotations [Gramann et al., 2018].

Heading computation is key for successful spatial orientation of humans and other animals. The registration of ongoing changes in the environment, perceived through an egocentric first-person perspective has to be integrated with allocentric, viewer-independent spatial information to allow complex navigation behaviors. The RSC provides the neural mechanisms to integrate egocentric and allocentric spatial information by providing an allocentric reference direction that contains the subject's current heading relative to the environment [Byrne et al., 2007]. Single-cell recordings in freely behaving animals have shown that the RSC is also implicated in heading computation [Sharp et al., 2001]. And although there is fMRI evidence that points to the same conclusion in humans that navigate in a virtual environment [Baumann and Mattingley, 2010], verifying this hypothesis in more naturalistic settings has remained elusive.

Recently, Gramann et al. [2018] used EEG synchronized to motion capture recordings combined with virtual reality (VR) to investigate the role of the RSC in heading computation of actively moving humans. Data were recorded from 19 participants using 157 active electrodes sampled at 1000 Hz and band-pass filtered from 0.016 Hz to 500 Hz using a BrainAmp Move System (Brain Products, Gilching, Germany). 129 electrodes were placed equidistant on the scalp and 28 were placed around the neck using a custom neckband. In that study, data from physically rotating participants were contrasted with rotations based on visual flow. In the physical rotation condition, participants wore a Vive

HTC head-mounted display (HTC Vive; $2 \times 1080 \times 1200$ resolution, 90 Hz refresh rate, 110° field of view). They were placed in a sparse VR environment devoid of any landmark information facing an orienting beacon at the beginning of each trial. The beacon was then replaced by a sphere that started rotating around them to the left or the right at a fixed distance with two different, randomly selected, velocity profiles on each trial. Participants were instructed to rotate on the spot to follow the sphere and keep it in the center of their visual field. The sphere movement was completed at an eccentricity randomly selected between 30° and 150° relative to the initial heading. When the sphere stopped, they had to rotate back and press a controller button to indicate when they believed to have reached their initial heading orientation. After the button press, the beacon would reappear and participants had to rotate to face the beacon and to start the next trial. In the joystick rotation condition, participants stood in front of a large TV screen (1.5 m viewing distance, HD resolution, 60 Hz refresh rate, 40″ diagonal size) controlling a gaming joystick to rotate in the same VR environment with an otherwise identical trial structure.

Using an ICA/dipole fitting approach, the data was analyzed with a focus on oscillatory activity of ICs located in or near the RSC. ICs were clustered using repetitive k-means clustering optimized to the RSC as the region of interest. Four subjects without an IC in the RSC were excluded from the analysis (21% of all participants). Subsequently, the wavelet (Morlet) time-frequency decomposition was computed for each IC in the RSC cluster for the rotation periods. The spectral baseline was defined as the 200 msec period before stimulus onset and subtracted from each time-frequency decomposition. To account for different trial durations, single trial time-frequency maps were linearly time-warped with respect to the presentation of the stimulus and rotation onset and offset to create time-warped event-related spectral perturbations (ERSPs). Using this approach, the data from the RSC cluster in the joystick rotation condition replicated previous studies using desktop navigation protocols and comparable data analysis approaches [Chiu et al., 2012; Gramann et al., 2010; Lin et al., 2015, 2018], exhibiting 1) a theta burst between stimulus

onset and movement onset and 2) alpha and beta desynchronization during the rotation. The physical rotation, however, had drastically different properties: no clear theta burst was present before movement onset, and only minor desynchronization in higher beta bands, but synchronization in the alpha and low beta bands after movement onset and delta and theta bands during the rotation (see Fig 2.10 **A-B**).

Here, we used the PEB+ algorithm to re-analyze the data. To this end, we further down-sampled the data to 250 Hz, removed the neck channels, applied a 0.5 Hz high-pass forward and backward FIR filter, and subtracted the common average reference. We co-registered each subject-specific 129-channels montage with the head surface of the "Colin27" template, computed each lead field matrix, and linearly warped the **A** dictionary to the space of the individualized template as explained in Section 2.3.1. Then we ran the PEB+ algorithm for each condition and computed the ERSPs of the centroid source activity (see Eq (2.15)) within the RSC. The computation of ERSPs was identical to the previous one, only the IC activity of the RSC cluster was replaced by PEB+ RSC source activity of all subjects.

Fig 2.10 **C-D** shows the PEB+ group ERSP for the joystick and physical rotation conditions as well as their difference. The top panel shows in red the location of the RSC in our template brain. Despite the differences between the two methodologies, our results largely replicate those in Gramann et al. [2018] displayed in panels **A-B**. A few differences between the two results are worth mentioning though. We point out that the differences in ERSP scales exhibited in panels **B** and **D** may be explained by different scales of the sources obtained by ICA and PEB+. Also, we note that the low-frequency power increase towards the end of the head rotation cycle in panel **D** Joystick condition can be explained by artifacts improperly removed near the end of a few trials. It should be emphasized that, unlike the approach used by Gramann et al. [2018], ours has the advantage of using data from all subjects without any cleaning in the time, channel, or trial domains, except for the inherent cleaning capabilities of the PEB+ algorithm. To increase the robustness to

residual artifacts, Fig 2.5 **E** suggests that a future research direction could explore the use of the log evidence yield by PEB+ to automatically downplay the influence of artifactual trials into post hoc statistical summaries.

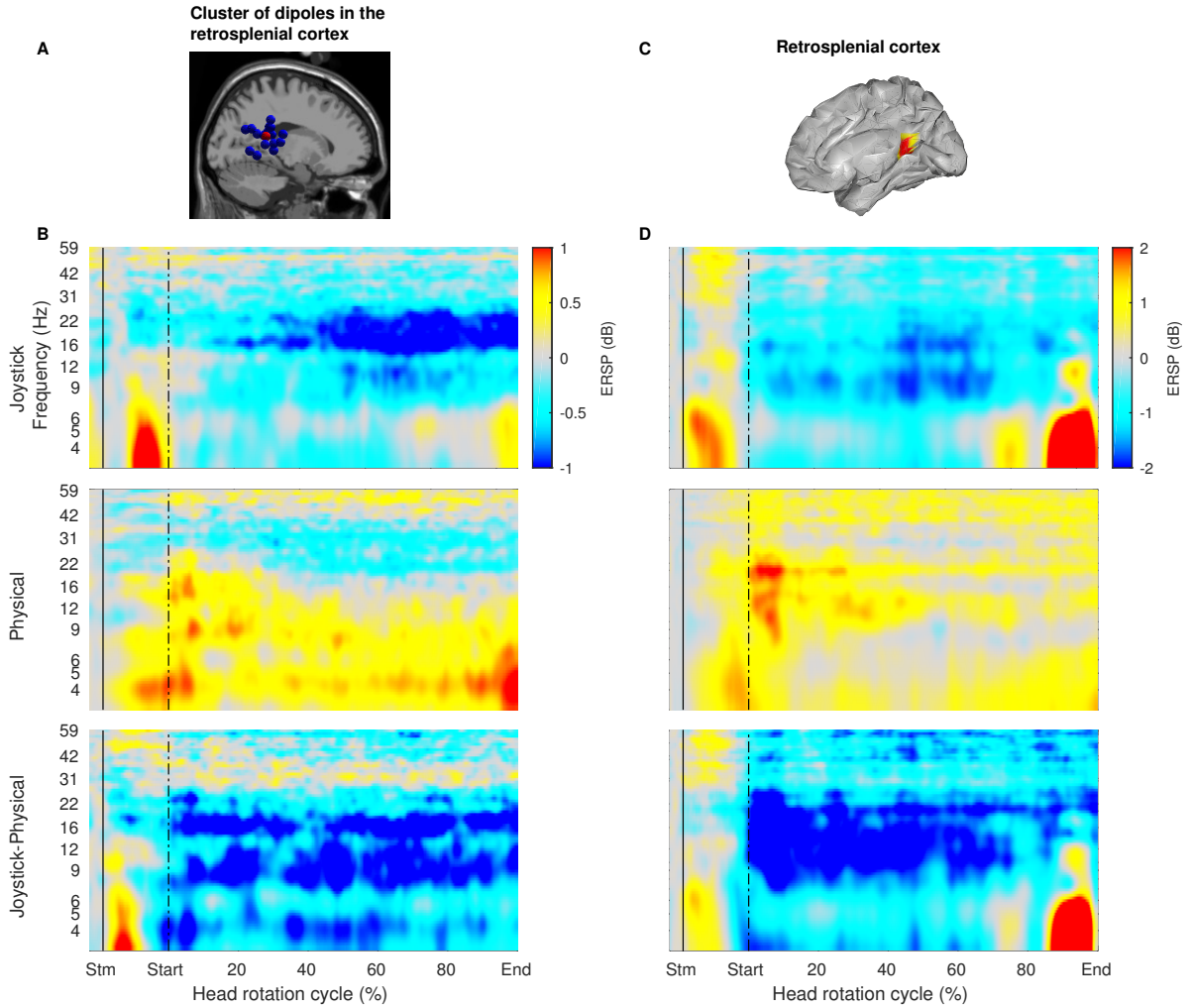


Figure 2.10. Event-related spectral perturbations (ERSPs) in the retrosplenial cortex (RSC). Panels **A** and **B** are adapted from [Gramann et al., 2018]. **A**: Cluster of IC equivalent current dipoles in or near the RSC. **B**: ICA derived ERSPs of the joystick and physical rotation conditions and their difference. **C**: Location of the RSC in the cortical surface of our template. **D**: PEB+ derived ERSPs of the joystick and physical rotation conditions and their difference. The x-axes at the bottom of panels **B** and **D** are annotated with the stimulus onset (Stm), movement onset (Start), percentage of the head rotation cycle, and movement offset (End).

2.4 Conclusions

In this chapter, we have extended the Parametric Empirical Bayes (PEB) framework previously proposed for electrophysiological source imaging [Henson et al., 2011; Wipf and Nagarajan, 2009] in two ways. First, we augmented the standard generative model of the EEG with a dictionary of artifact scalp projections obtained empirically. In our model, we captured EOG, EMG, and single-channel spike artifacts. Second, we used an anatomical atlas to parametrize a source prior that encourages sparsity in the number of active cortical regions, which has the desired property of inducing the segregation of the cortical electrical activity into a few maximally independent components with known anatomical support. We used these elements to develop the PEB+ inversion algorithm. Under the proposed framework, dissimilar problems such as data cleaning, source separation, and imaging can be understood and solved in a principled manner using a single algorithm. Furthermore, we used our framework to point out the connections between distributed source imaging and Independent Component Analysis (ICA), two of the most popular approaches for EEG analysis that are often perceived to be at odds with one another.

We used publicly available data from two independent studies to develop and test the proposed algorithm. In particular, we have shown that PEB+: 1) outperforms classic PEB for source imaging when artifacts are present in the data, while on clean data their performance is comparable, 2) outperforms Infomax ICA for source separation on short blocks of data, thereby showing potential for tracking non-stationary cortical dynamics, and 3) unlike the popular Artifact Subspace Removal algorithm, it can reduce artifacts without significantly distorting epochs of clean data. Furthermore, we were able to replicate the main finding of a study that looked into the dynamics of the retrosplenial cortex (RSC) supporting heading computation during full-body rotations.

The ability to estimate the time series of EEG sources that correspond to known anatomical locations accounting for the influence of artifacts without user intervention, as

well as its online adaptation, makes the PEB+ algorithm appealing for established ERP paradigms as well as MoBI. We believe that the proposed algorithm can help to solve basic research questions employing EEG as the functional imaging modality, and at the same time constitute a biologically-grounded signal processing tool that can be useful to translational efforts.

2.5 Acknowledgements

This research was supported by NIMH Training Fellowship in Cognitive Neuroscience (AO), UC San Diego Chancellor's Research Excellence Scholarship (JM, AO), and UC San Diego School of Medicine start-up funds (JM). The PEB+ algorithm is copyrighted for commercial use (UC San Diego Copyright #SD2019-810) and free for research and educational purposes.

In part, this chapter has been submitted for publication, and it is available as a preprint in Ojeda, A., Klug, M., Kreutz-Delgado, K., Gramann, K., & Mishra, J. (2019). A Bayesian framework for unifying data cleaning, source separation and imaging of electroencephalographic signals. https://doi.org/10.1101/559450. I was the primary investigator and author of this paper.

Chapter 3

Fast and Robust Block-Sparse Bayesian Learning for EEG Source Imaging

Abstract

We propose a new Sparse Bayesian Learning (SBL) algorithm that can deliver fast, block-sparse, and robust solutions to the EEG source imaging (ESI) problem in the presence of noisy measurements. Current implementations of the SBL framework are computationally expensive and typically handle fluctuations in the measurement noise using different heuristics that are unsuitable for real-time imaging applications. We address these shortcomings by decoupling the estimation of the sensor noise covariance and the sparsity profile of the sources, thereby yielding an efficient two-stage algorithm. In the first stage, we optimize a simplified non-sparse generative model to get an estimate of the sensor noise covariance and a good initialization of the group-sparsity profile of the sources. Sources obtained at this stage are equivalent to those estimated with the popular inverse method LORETA. In the second stage, we apply a fast SBL algorithm with the noise covariance fixed to the value obtained in the first stage to efficiently shrink to zero groups of sources that are irrelevant for explaining the EEG measurements. In addition, we derive an initialization to the first stage of the algorithm that is optimal in the least squares sense, which prevents delays due to suboptimal initial conditions. We validate our method on both simulated and real EEG data. Simulations show that the method is

robust to measurement noise and performs well in real-time, with faster performance than two state of the art SBL solvers. On real error-related negativity EEG data, we obtain source images in agreement with the experimental literature. The method shows promise for real-time neuroimaging and brain-computer interface applications.

3.1 Introduction

A fundamental issue that prevents electroencephalography (EEG) from becoming a more widely used imaging modality for studying the functioning brain is that the mapping from scalp sensors to brain sources is not unique [Michel and Murray, 2012]. As infinite configurations of currents in the brain can elicit the same EEG scalp topography, EEG source imaging (ESI) entails solving an ill-posed inverse problem [Lopes da Silva, 2013]. Compared to other brain imaging modalities such as functional magnetic resonance imaging (fMRI), positron emission tomography (PET), or functional near-infrared spectroscopy (fNIRS), the temporal resolution of EEG allows the study of brain dynamics at their natural time scale. However, EEG source estimates only become available after solving the inverse problem. Although this is not a handicap when the analysis can be done offline, there are many applications of interest where real-time source estimates are desirable. For example, in applications involving brain-computer interfaces[1] (BCI) and neurofeedback, brain responses at specific anatomical locations can be used to improve real-time predictions of subject's intentions compared to sensor-based predictions [Bauer and Pllana, 2014; Edelman et al., 2016; Haufe et al., 2011; Noirhomme et al., 2008]. In this chapter, we focus on a Bayesian inversion scheme that facilitates the development of such real-time imaging applications.

As we mentioned in Chapter 2, we can deal with the ill-posed nature of the inverse problem by imposing regularizing constraints to "encourage" a solution to belong to a particular functional space with biological relevance [Tikhonov and Arsenin, 1978]. There

[1] Also known as brain-machine interfaces (BMI).

are several software packages readily available to tackle this problem within the framework of constraint optimization [Boyd et al., 2011; Combettes and Pesquet, 2011; Tomioka et al., 2009; Vega-Hernández et al., 2008]. Most of these packages are used in combination with different heuristics to determine the amount of regularization needed at each time step to optimize the value of $\mathbf{g}_k$ [Bauer and Lukas, 2011; Golub et al., 1979; Reichel and Sadok, 2008]. Selecting the right amount of regularization, however, is a crucial issue because, if set suboptimally, even small perturbations in the data can lead to neurophysiologically unreasonable results [Bertero et al., 1988].

One way of estimating EEG sources subject to biological and mathematical constraints in a principled manner is to use a probabilistic approach known as Parametric Empirical Bayes. PEB has its roots in the statistical literature starting in the 1950's with the work of Herbert Robbins [Robbins, 1956] and further developed by Bradley Efron, Carl Morris, and others [Casella, 1985; Efron and Morris, 1972a,b; Morris, 1983], exploiting the idea of using data to infer the prior distribution of a statistical model in contrast to the more conventional Bayesian method of fixing the prior before any data are observed. In the context of sparsity-inducing priors, PEB is sometimes referred to as Sparse Bayesian Learning (SBL) [Tipping, 2001]. This approach has been used to invert hierarchical models of fMRI and PET [Friston et al., 2002] as well as EEG responses [Henson et al., 2011; Phillips et al., 2005]. It has been shown that PEB offers a unifying framework in which a variety of popular EEG inverse solvers can be obtained [Wipf and Nagarajan, 2009]. In this framework, sparse source estimates are obtained by enforcing appropriate priors at two levels of inference [MacKay, 2008b]. At the first level, a prior over the source vector is controlled by a set of hyperparameters, each of which represents the variance (alternatively the precision, i.e., inverse variance) of each source. At the second level, sparsity-inducing hyperpriors are used so that when the hyperparameters are learned from the data, most of them go to zero (or infinity in the case of precision hyperparameters), a property known as *Automatic Relevance Determination* (ARD) [MacKay, 1992; Neal, 1996]. For linear

models such as the one introduced in Chapter 2, once the sparsity profile is determined, the maximum a posteriori (MAP) source estimate, $\hat{\mathbf{g}}_k$, can be easily computed in closed form [Bishop, 2006].

In this chapter, we use "SBL" rather than "PEB" to be consistent with the terminology of Wipf and Nagarajan [2009], Zhang and Rao [2013] , and our own publication [Ojeda et al., 2018], heavily cited here. We point out, however, that in the context of sparse EEG source imaging, both frameworks are equivalent as they rest on the same general sparsity-inducing priors, hyperpriors, and the optimization of covariance components.

Assuming that the statistics of the noise affecting the EEG samples can be learned from pre-stimulus data, sparse sources can be estimated efficiently using the Champagne algorithm [Wipf et al., 2010]. In many practical BCI applications, however, the EEG data are not time-locked to any specific event and the signal to noise ratio (SNR) can be subject to random fluctuations due to a variety of physiological and environmental conditions [Costa et al., 2016; Kilicarslan et al., 2016], and hence there is a need for inversion algorithms with built-in online noise adaptation. Some authors have proposed combining noise adaptation and sparse source estimation in a double optimization algorithm [Zhang and Rao, 2013]. We will show in Section 3.3.4 that such an approach may incur a significant computational cost that limits the usability of the SBL framework for real-time ESI.

In the context of EEG source estimation, the use of global sparsity and local correlation (smoothness) constraint can be motivated by the fact that EEG signals are due to the electrical activity of neural masses locally synchronized in space and time. Block-SBL (BSBL) approaches comprise different flavors of the SBL framework in which the sparsity constraints can be enforced at the level of groups of sources [Zhang and Rao, 2011, 2013], thereby providing a principled way to model intra-group source correlations. A similar approach termed Multiple Sparse Priors (MSP) has been developed independently by [Friston et al., 2008]. In BSBL, block-sparse source estimates that are easy to interpret can be induced by partitioning the source space according to an atlas of the cerebral

cortex [Saha et al., 2017]. In this chapter, we build on this approach. In addition to its biological motivation, group-sparsity is appealing for real-time ESI because the reduced number of hyperparameters that need to be learned from the data can help to further accelerate the algorithm. The objective of this chapter is to develop a BSBL algorithm with built-in noise adaptation that is suitable for real-time imaging applications. The motivation for our method is an approximation to the model evidence proposed by [Friston and Penny, 2011] that allows for decoupling the estimation of the sensor noise covariance from the optimization of the ARD priors. This yields a fast and robust two-stage BSBL algorithm which we refer to as BSBL-2S. In the first stage, we optimize a simplified non-sparse generative model to get an estimate of the sensor noise covariance and a good initialization of the group-sparsity profile of the sources. In the second stage, we apply a fast SBL algorithm with the noise covariance fixed to the value obtained in stage one to efficiently shrink to zero groups of sources that are irrelevant for explaining the observed EEG data. Furthermore, we derive an initialization to the first stage of the algorithm that is optimal in the least squares sense, thereby preventing delays due to suboptimal initial conditions. To achieve globally sparse and locally smooth source estimates while reducing the computational cost of our algorithm, we use an anatomical atlas of the cortex and discrete Laplacian operators to furnish the source prior of our generative model.

The remainder of the chapter is organized as follows. In Sections 3.2.1-3.2.6 we develop the BSBL-2S algorithm. In Sections 3.3.1-3.3.6 we validate our method on simulated data. In Section 3.3.7 we show an application on real EEG data for imaging the sources of error-related negative/positive scalp potentials. We finalize in Section 3.4 with our concluding remarks.

3.2 Methods

As we mentioned in Chapter 2, in practice, raw EEG data are affected by artifacts of different nature such as interference from the 50/60 Hz AC line, pseudo-random muscle, and movement artifacts, among others. However, to simplify the exposition, in this chapter we don't model artifacts explicitly as in the PEB+ framework proposed in Chapter 2. Instead, we assume that before source estimation, EEG samples pass through an online-capable artifact rejection procedure where high-variance and non-Gaussian artifacts are largely removed [Kilicarslan et al., 2016; Matiko et al., 2013; Mullen et al., 2015]. We do this so that later in the chapter we can benchmark our algorithm against others without built-in artifact modeling on equal footing.

To invert the probabilistic generative model of Chapter 2 we maximize the evidence $p(\mathbf{Y}|\mathcal{M})$ with respect to the hyperparameters (see Section 2.2.2). We note that for the purpose of deriving an algorithm to optimize the evidence, the simplifying assumption that $\mathbf{A} \in \emptyset$, does not constitute a loss of generality. Due to the linearity of our generative model, considering a nonempty artifact dictionary implies that we just need to recast the algorithm developed here (see pseudo-code 1) and detailed in appendix C with $\mathbf{L} \leftarrow \mathbf{H}$, $\mathbf{\Sigma}_g \leftarrow \mathbf{\Sigma}_x$, so that we obtain $\hat{\mathbf{X}}_k$ estimates instead of $\hat{\mathbf{G}}_k$.

As in Chapter 2, we define the prior source covariance in terms of the following covariance components Wipf and Nagarajan [2008],

$$\mathbf{\Sigma}_g(\gamma) = \sum_{i=1}^{N_g} \gamma_i \mathbf{C}_i, \quad i = 1, \ldots, N_g \tag{3.1}$$

where $\mathbf{C}_i \in \mathbb{R}^{N_g \times N_g}$ represents a known nonoverlapping intra-group pseudo-covariance matrix and $\gamma \in \mathbb{R}^{N_g}$ is a nonnegative scale vector. $\mathbf{C}_i$ is zero everywhere except for a block in the main diagonal that takes the value $(\Delta_i^T \Delta_i)^{-1}$, where the matrix Δ_i is a discrete Laplacian operator which allows enforcing spatial smoothness within each group.

To obtain each Δ_i, we first build a discrete Laplacian operator Δ defined in the entire cortical mesh [Oostendorp et al., 1989], excluding the vertices that belong to the corpus callosum, then we zero out the columns of Δ that are not in the ith group. We partition the source space into N_g regions of interest (ROIs) according to the Desikan & Killiany atlas [Desikan et al., 2006], as depicted in Fig. 3.1. Within the framework outlined in this section, different brain parcellations can be included based on structural and functional data contributed by other neuroimaging modalities [Lei et al., 2015]. It has been shown that the parametrization of Eq. (3.1) allows for recovery of a variety of physiologically relevant source configurations, from broadly smooth to extremely sparse [Friston et al., 2008].

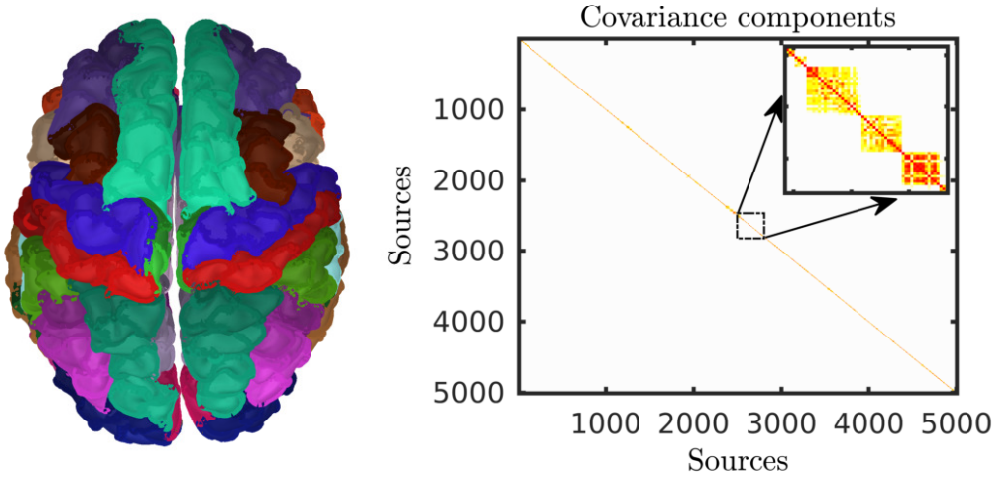


Figure 3.1. Definition of covariance components. The left panel shows the Desikan & Killiany atlas used in this chapter subdivided into $N_g = 488$ regions of interest (ROI), each covering approximately 5 cm^2 of surface area. Each color represents a meta-ROI which is subdivided into smaller regions as indicated by the boundary marks. The right panel shows the unweighted sum of covariance components $\mathbf{C} = \sum_{i=1}^{N_g} \mathbf{C}_i$, where each $\mathbf{C}_i$ matrix takes values according to the neighborhood structure of the correspondent ith ROI.

The combination of discrete Laplacian operators and a cortical atlas has been used before to constraint EEG sources [Trujillo-Barreto et al., 2004]. In the context of BCI, Duque-Munoz et al. [2016] recently proposed to reduce computational cost by using the

Broadmann atlas to drastically limit the source space to a few exemplars inside each ROI. A key distinction between our approach and the latter is that we do not use ROIs to reduce the source space at the outset, which would reduce the spatial resolution of our method. Instead, we use ROIs to induce sparsity at the level of groups of sources, i.e., in the first level of inference we still estimate the amplitude of N_g sources.

3.2.1 Two-stage evidence optimization (BSBL-2S)

The motivation for our method is an approximation to the model evidence proposed by Friston and Penny [2011] that allows for decoupling the estimation of the sensor noise variance λ from the optimization of the sparsity profile of the sources encoded by γ. Due to the ARD properties of our generative model, in the process of learning γ most entries shrink to zero, causing the corresponding sources to have zero variance. This allows for removal of the respective columns of $\mathbf{L}$ from the model of Eq. (2.1), thereby reducing the column dimension of $\mathbf{L}$. We refer to models where some of its sources have identically zero variance as *reduced* (sparse) models, $\mathcal{M}_R \Leftrightarrow \gamma \succeq 0$. Note that if a source has zero mean and variance, it is identically zero almost surely and contributes nothing to the value of $\mathbf{Y}$. Conversely, we say that a model is *full* (non-sparse) if all the sources can have finite non-zero variance, $\mathcal{M}_F \Leftrightarrow \gamma \succ 0$.

Friston and Penny [2011] showed that if the likelihoods of the reduced and full models are assumed to be the same (i.e. models $\mathcal{M}_F$ and $\mathcal{M}_R$ yield the same reconstruction error), the inversion of the full model can be used as a stepping stone to invert a reduced one. Their so-called post hoc model selection technique yields simple expressions for the reduced model evidence and source posterior, but requires matrix-vector operations in the source space $\mathbb{R}^{N_g}$ which are unsuitable for real-time imaging applications of medium to high spatial resolution. Nonetheless, their result is relevant for obtaining a fast and noise-adaptive SBL algorithm because a constrained parameterization of the full model may facilitate the optimization of the sensor noise variance λ with respect to $p(\mathbf{Y}|\mathcal{M}_F)$,

clearing the way for the subsequent application of an efficient learning rule for γ. We develop this idea further below.

Let $\mathcal{M}_F = \{\lambda, \gamma \in \Gamma\}$ be our chosen class of "γ-constrained" full models, where Γ denotes the following constraint set,

$$\Gamma = \{\gamma_i = \gamma_F | \gamma_F > 0, i = 1, \ldots, N_g\} \tag{3.2}$$

We can learn $\mathcal{M}_F$ by finding the mode of the full model evidence as follows,

$$\hat{\mathcal{M}}_F = \arg\max_{\lambda, \gamma_F} p(\mathbf{Y}|\lambda, \gamma_F) \tag{3.3}$$

Eq. (3.3) defines our "stage one" optimization problem. Conveniently, the constraint (3.2) reduces problem (3.3) to the learning of just two scalar hyperparameters (λ, γ_F) which, in the context of EEG where N_g can be quite large (see Fig. 3.1), is a significant reduction in complexity. Furthermore, it forces the source prior covariance defined in Eq. (3.1) to take the simpler form

$$\mathbf{\Sigma}_g(\gamma) = \gamma_F \mathbf{C} \tag{3.4}$$

where $\mathbf{C} \triangleq \sum_i \mathbf{C}_i$ is known. Eq. (3.4) yields the smoothness prior used in the popular source estimation method LORETA [Pascual-Marqui et al., 1994]. LORETA has been proposed as a valid choice for implementing real-time (non-sparse) brain imaging pipelines [Courellis et al., 2017; Mullen et al., 2015; Pieloth et al., 2014; Pion-Tonachini et al., 2015], thus, the solution of Eq. (3.3) can yield a reasonable starting point for a subsequent sparsity-inducing pruning algorithm, as we show is indeed the case in Section 3.3.2.

The "stage two" of our algorithm is SBL, used for the pruning of redundant sources from the γ-constrained optimal full model that was found in the previous stage,

$$\hat{\mathcal{M}}_R = \arg\max_{\gamma} p(\mathbf{Y}|\hat{\lambda}, \gamma) \tag{3.5}$$

which we solve using the so-called γ-MAP algorithm [Wipf and Nagarajan, 2008]. We note that by using γ-MAP, our stage two is equivalent to the Champagne algorithm [Wipf et al., 2010]. Our method differs from Champagne in that the latter estimates the noise covariance Σ_e from pre-stimulus data while ours do not require time-locking to any experimental event. Next, we summarize the proposed algorithm.

3.2.2 Algorithm outline

Algorithm 1 gives a high-level pseudocode outline of the two-stage method developed so far (see a detailed pseudocode in C). The algorithm takes as input a block of EEG samples $\mathbf{Y}$ and returns the correspondent MAP source estimates in the matrix $\hat{\mathbf{G}}$. In line 2 we obtain an estimate of the full model $\hat{\mathcal{M}}_F = \{\hat{\lambda}, \hat{\gamma}_F\}$ by solving problem (3.3). In line 3 we use $\hat{\gamma}_F$ as the initial condition for the solution of problem (3.5). In line 5 we fix λ to the value $\hat{\lambda}$ and continue the optimization of the evidence, now allowing irrelevant groups of sources to shrink to zero. Finally, in line 7 we use the optimal sparse model $\hat{\mathcal{M}}_R = \{\hat{\lambda}, \hat{\gamma}\}$ to estimate the sources $\hat{\mathbf{G}}$.

We present simulations showing that this two-stage strategy works well in Section 3.3.4. But first, we turn to developing the details of the algorithms needed to perform the optimizations (3.3) and (3.5).

Algorithm 1. Two-stage evidence optimization

Input: Y
Output: $\hat{\mathbf{g}}$
1: **stage one (LORETA):**
2: $\hat{\lambda}, \hat{\gamma}_F \leftarrow \arg\max_{\lambda,\gamma_F} p(\mathbf{Y}|\lambda, \gamma_F)$
3: $\gamma_i \leftarrow \hat{\gamma}_F,\ i = 1, \ldots, N_g$
4: **stage two (pruning):**
5: $\hat{\gamma} \leftarrow \arg\max_{\gamma} p(\mathbf{Y}|\hat{\lambda}, \gamma)$
6: Source estimation:
7: $\hat{\mathbf{G}} \leftarrow \mathbf{\Sigma}_g(\hat{\gamma})\mathbf{L}^T\mathbf{\Sigma}_y(\hat{\lambda}, \hat{\gamma})^{-1}Y$

3.2.3 Optimization of the model log evidence

The maximization of the model evidence is equivalent to minimizing the so-called type-II Maximum Likelihood (ML-II) cost function Barber [2012], which is also known as the evidence framework MacKay [2008b]. The only implication of using ML-II for Algorithm 1 is the switch in lines 2 and 5 to the minimization of the functional (3.6) subject to the appropriate constraint. Applying $-2\log(\cdot)$ to Eq. (2.19) yields,

$$\mathcal{L}(\mathcal{M}) = \underbrace{\log|\boldsymbol{\Sigma}_y|}_{\text{Complexity}} + \underbrace{\text{trace}\big(\mathbf{C}_y\boldsymbol{\Sigma}_y^{-1}\big)}_{\text{Accuracy}} \tag{3.6}$$

where $\mathbf{C}_y = N^{-1}\mathbf{Y}\mathbf{Y}^T$ is the empirical data covariance. The first term in Eq. (3.6) measures model complexity. Geometrically, this term represents the volume of an ellipsoid defined by $\boldsymbol{\Sigma}_y$; as the axes of the ellipsoid shrink, the volume is reduced. The second term measures model accuracy; i.e., how similar are the empirical and analytic covariances $\mathbf{C}_y$ and $\boldsymbol{\Sigma}_y$.

Although several learning rules for λ and a scalar γ based on the evidence framework are readily available MacKay [1992]; Trujillo-Barreto et al. [2004], they require vector operations in $\mathbb{R}^{N_g}$ which for high-resolution source spaces could hamper real-time performance. Therefore we next develop learning rules equivalent to those traditionally used in the evidence framework which only require operations in the space of the data $\mathbb{R}^{N_y}$ (see details in A).

3.2.4 Stage one: optimization of the full model

Inserting the γ-constrained source covariance (3.4) in Eq. (2.11) (evaluated in $\mathbf{L}$ since here we ignore artifacts) yields the simplified analytic data covariance,

$$\boldsymbol{\Sigma}_y(\lambda, \gamma_F) = \lambda I_{N_y} + \gamma_F \mathbf{L}\mathbf{C}\mathbf{L}^T \tag{3.7}$$

Then, we use the relation $\mathbf{C} = (\Delta^T \Delta)^{-1}$, where Δ is the discrete surface Laplacian operator introduced in Section 3.2, and the singular value decomposition of the standardized lead field matrix $\mathbf{L}\Delta^{-1} = \mathbf{U}\text{diag}(s_i)\mathbf{V}^T$ to succinctly write $\mathcal{L}(\mathcal{M}_F)$ as a spectral function (a function that depends only on the eigenvalues of a matrix),

$$\mathcal{L}(\mathcal{M}_F) = \underbrace{\sum_{i=1}^{N_y} \log \psi_i}_{\log|\mathbf{\Sigma}_y|} + \underbrace{\sum_{i=1}^{N_y} \tilde{y}_i^2 \psi_i^{-1}}_{\text{trace}(C_y \mathbf{\Sigma}_y^{-1})} \tag{3.8}$$

with

$$\begin{aligned} \psi_i &= \gamma_F s_i^2 + \lambda \\ \tilde{y}_i^2 &= N^{-1} \|\tilde{\mathbf{Y}}_i\|_2^2, \quad i = 1, \ldots, N_y \end{aligned} \tag{3.9}$$

where ψ_i are the regularized eigenvalues of Eq. (3.7) and $\tilde{\mathbf{Y}}_i$ is the ith row of the matrix $\tilde{\mathbf{Y}} = \mathbf{U}^T \mathbf{Y}$.

We obtain the optimal hyperparameter estimates $\hat{\lambda}$ and $\hat{\gamma}_F$ simply by finding the root of the gradient of $\mathcal{L}(\mathcal{M}_F)$, which using Eq. (3.8) yields the following fixed-point update rules,

$$\hat{\lambda} \leftarrow \hat{\lambda} \frac{\sum_{i=1}^{N_y} \tilde{y}_i^2 \psi_i^{-2}}{\sum_{i=1}^{N_y} \psi_i^{-1}}, \quad \hat{\gamma_F} \leftarrow \hat{\gamma_F} \frac{\sum_{i=1}^{N_y} \tilde{y}_i^2 s_i^2 \psi_i^{-2}}{\sum_{i=1}^{N_y} s_i^2 \psi_i^{-1}} \tag{3.10}$$

3.2.5 Stage one least squares initialization

Initializing Eq. (3.10) at random could reduce the convergence speed of stage one depending on how far are the initial points λ_0 and γ_0 in the ML-II landscape from the optimal $\hat{\lambda}$ and $\hat{\gamma}_F$ (see Fig. 3.3). In this chapter, we propose using a starting point that is optimal in the least squares (LS) sense.

Inspecting the trace term in Eq. (3.8) we see that if the model fits the data perfectly the following condition is met,

$$\sum_{i=1}^{N_y} \tilde{y}_i^2 \psi_i^{-1} = N_y \tag{3.11}$$

yielding the following system of linear equations,

$$\underbrace{\begin{bmatrix} s_1^2 & 1 \\ \vdots & \vdots \\ s_{N_y}^2 & 1 \end{bmatrix}}_{\mathbf{S}} \begin{bmatrix} \gamma_0 \\ \lambda_0 \end{bmatrix} = \underbrace{\begin{bmatrix} \tilde{y}_1^2 \\ \vdots \\ \tilde{y}_{N_y}^2 \end{bmatrix}}_{\tilde{\mathbf{y}}^2} \tag{3.12}$$

which we can solve analytically by least squares as,

$$\begin{bmatrix} \gamma_0 \\ \lambda_0 \end{bmatrix} = (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\tilde{\mathbf{y}}^2 \tag{3.13}$$

Note that the LS estimator of Eq. (3.13) is prone to overfitting, however, a few iterations of the update rules in Eq. (3.10) will cause the complexity term to be taken into account thereby moving the hyperparameter set towards more plausible values. The properties of the estimator in Eq. (3.13) are not discussed further in this chapter.

3.2.6 Stage two: pruning of irrelevant sources

In this section we perform stage two of Algorithm 1 by leveraging the γ-MAP learning rule, initialized by the γ-constrained solution $\hat{\mathcal{M}}_F$. Let $\{\mathbf{L}_1, \ldots, \mathbf{L}_{N_g}\}$ be a set of standardized observation operators such that $\mathbf{L}_i \triangleq \mathbf{L}\Delta_i^{-1}$, where Δ_i are the ROI-specific Laplacian operators introduced in Section 3.2. Note that the $\mathbf{L}_i$ matrices can be significantly reduced by removing the columns corresponding to the zero entries of Δ_i^{-1}, thereby saving compute time and storage. Then, we update each γ_i as follows Wipf and Nagarajan [2008],

$$\hat{\gamma}_i \leftarrow \frac{\hat{\gamma}_i}{\sqrt{N}} \left\| \mathbf{L}_i^T(\boldsymbol{\Sigma}_y)^{-1}\mathbf{Y} \right\|_{\mathcal{F}} \left(\text{trace}\left[\mathbf{L}_i^T\boldsymbol{\Sigma}_y^{-1}\mathbf{L}_i \right] \right)^{-1/2} \tag{3.14}$$

where $\mathbf{\Sigma}_y$ is updated at each iteration as,

$$\mathbf{\Sigma}_y \leftarrow \hat{\lambda}\mathbf{I}_{N_y} + \sum_{i=1}^{N_g} \hat{\gamma}_i \mathbf{L}_i \mathbf{L}_i^T \tag{3.15}$$

The complexity of learning rule (3.14) is at most $\mathcal{O}(N_y^3 N_g)$, but with access to parallel hardware each γ_i could be optimized simultaneously, thereby lowering the compute cost even further. Although we do not use such implementation here, we show in Section 3.3.4 that real-time performance is attainable.

3.3 Results and Discussion

In this section we first use simulated data to illustrate different aspects of the BSBL-2S algorithm and test speed performance and noise adaptation. In addition to the log evidence score, we use the area under the convex-hull (AUC) of the ROC curve [Provost and Fawcett, 2000] to assess the reconstruction accuracy with respect to the ground truth source maps. We calculate the AUC compensating against class imbalance as done in previous neuroimaging studies [Chowdhury et al., 2013; Trujillo-Barreto et al., 2008]. Second, we used BSBL-2S for imaging the sources of error-related negative/positive scalp potentials obtained from real EEG data.

3.3.1 Data simulation

We constructed simulations based on a montage of $N_y = 64$ sensors (a superset of the standard 10-20 system) and $N_g = 5003$ sources distributed over the cortical mesh. The source orientations were constrained to be normal to the cortical surface. The lead field matrix $\mathbf{L}$ was computed from a four-layer (scalp, outer skull, inner skull, and cortex) head model derived from the MRI template “Colin27" [Holmes et al., 1998] using the boundary element method solver OpenMEEG [Gramfort et al., 2010]. We simulated the ground truth source vector $\mathbf{g}_{true}$ by placing a truncated Gaussian at the center of four

ROIs (middle frontal gyrus (R), caudal middle frontal gyrus (L), inferior parietal cortex (L/R)), each one covering an area of approximately 5 cm^2 (see Fig. 3.2). We generated synthetic EEG samples at 250 Hz by adding iid Gaussian noise to the ground truth source projection,

$$\mathbf{y}_k = \mathbf{L}\mathbf{g}_{true} + \mathbf{e}_k, \quad k = 1, \ldots, N \tag{3.16}$$

The block smoothness and sparsity constraints were constructed based on the 488-ROI sub-parcellation of the Desikan & Killiany atlas shown in Fig. 3.1, which we obtained with the help of the Brainstorm software [Tadel et al., 2011]. The discrete surface laplacian operator Δ was extracted from the irregular cortical mesh using MATLAB code by Darren Weber and Robert Oostenveld that implemented the formulas in [Oostendorp et al., 1989]. The discrete Laplacian operators Δ_i were constructed zeroing out the rows of Δ that do not correspond to the ith ROI.

We define the SNR used in our experiments as the power of Hx_{true} divided by the average power of $\mathbf{e}_k$. Likewise, the AUC is calculated between the average estimated PCD, $\hat{\mathbf{g}}_k^2$, and the ground truth PCD, x_{true}^2. We define the convergence of the ML-II cost function as a difference in the log evidence of two consecutive iterations of the optimization algorithm equal to 0.1[2]. All the simulations were implemented in MATLAB and ran on a 4-core i5 Intel CPU laptop with 8 GB of RAM.

3.3.2 BSBL-2S model inversion

In this experiment, we illustrate how ESI is realized through the BSBL-2S algorithm. The objective here is to show qualitatively and quantitatively how the block-sparse source vectors $\mathbf{g}_k$ are obtained from the simulated noisy measurements $\mathbf{y}_k$ described in Section 3.3.1. Fig. 3.2 shows images obtained at each stage of the algorithm. To better illustrate the difference between each type of solution, the source maps shown correspond to the

[2]We note that a difference of 0.1 in log evidence units implies that the models compared in iterations i and $i+1$ have almost the same probability, $1.052 \cdot p(\mathbf{Y}|\mathcal{M}_i) \approx p(\mathbf{Y}|\mathcal{M}_{i+1})$.

power $\hat{\mathbf{g}}_k^2$ averaged over $N = 10$ samples, i.e. such that spurious positive and negative sources on each trial cannot cancel each other out, thereby artificially improving the results.

As shown in Fig. 3.2, the full model $\hat{\mathcal{M}}_F$, estimated during the first stage of the algorithm, converged in seven iterations. Regarding reconstruction accuracy, the non-sparse solution achieved an AUC of 0.8818. However, as it is typical with LORETA-type images, some undesired low-amplitude ghost sources remained active. The algorithm proceeds to the second stage (pruning) using the update rule (3.14) with λ fixed to the previously learned value $\hat{\lambda}$. The reduced model $\hat{\mathcal{M}}_R$ converged in an additional seven iterations. Qualitatively, we see that ghost sources were mostly removed from the sparse solution, achieving an AUC of 0.9504. Note that AUC increases with the log evidence of the model as the estimated source map approaches the ground truth. In addition, we compared models $\mathcal{M}_F$ and $\mathcal{M}_R$ according to their log Bayes factor, $B_{R,F} = \log p(\mathbf{Y}|\hat{\mathcal{M}}_R) - \log p(\mathbf{Y}|\hat{\mathcal{M}}_F)$, where $B_{R,F} \approx 15$ indicates that the data are more likely to be generated by the sparse model $\mathcal{M}_R$ [Kass and Raftery, 1995].

3.3.3 Effect of the initial conditions

In this section we investigate, empirically, how random initial conditions affect the convergence properties of our algorithm. We simulate 1000 values for λ_0 and γ_0 by drawing independent samples from a uniform distribution in a log-space,

$$\lambda_0, \gamma_0 \sim U\left[10^{-6}, 10^{6}\right] \tag{3.17}$$

The sampling intervals were set to cover values that yield physiologically plausible solutions, i.e., avoiding source vectors of all zeros or infinity.

Starting from each of the 1000 initial conditions, we optimized λ and $N_g = 488$ γ_i hyperparameters. Fig. (3.3) plots the log evidence (denoted by the colors) at each

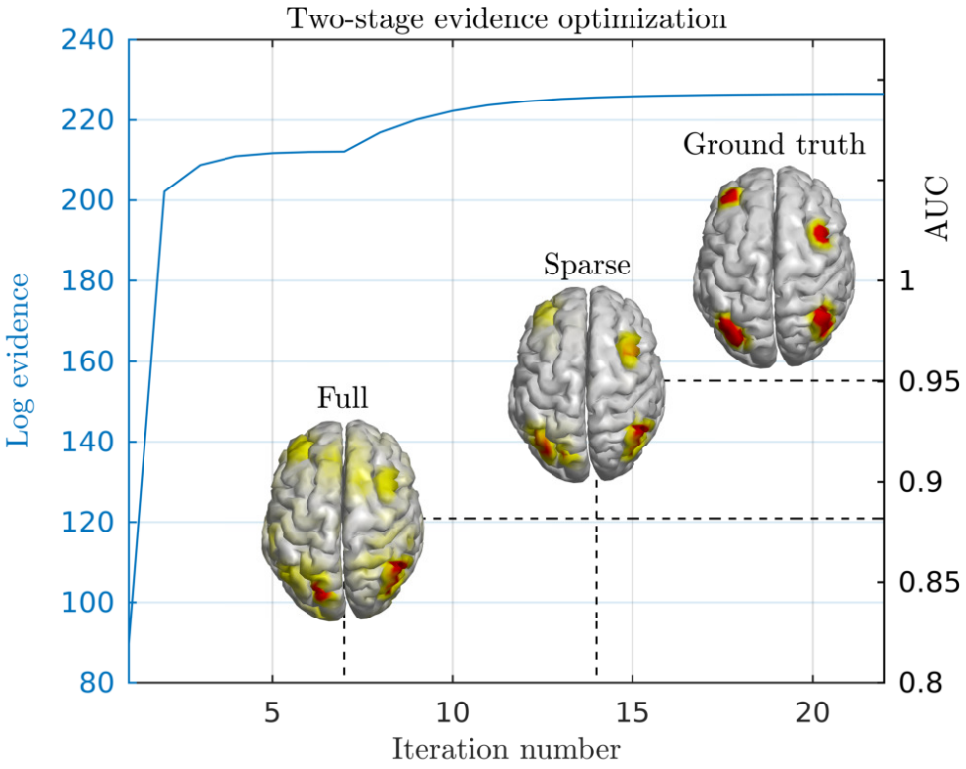


Figure 3.2. Two-stage log evidence optimization. The blue trace represents the optimization of the log evidence (arrived at via the iterative minimization of $\mathcal{L}$). The first 7 iterations are spent in the optimization of the full model, after that, an increase in log evidence is achieved by removing unnecessary sources. The 3D maps represent the average source power correspondent to the full, sparse, and ground truth maps. The right y-axis shows the AUC at each stage. The SNR of the simulation was set to 4 dB.

optimization iteration. The simulations (bottom axis) were sorted by the number of iterations required for stage one of the algorithm to converge. We note that, regardless of the initial conditions, the two stages of the algorithm appear to converge to approximately the same value of log evidence. To test this, we computed the Bayes factor between the models with maximum and minimum log evidence, $\log p(\mathbf{Y}|\mathcal{M}_R^+) = 135.968$ and $\log p(\mathbf{Y}|\mathcal{M}_R^-) = 135.5518$ respectively, yielding $B_{R^+,R^-} = 0.411$. Since $2B_{R^+,R^-} < 2$, we can discard the hypothesis that the 1000 final models are distinguishable from one another [Kass and Raftery, 1995].

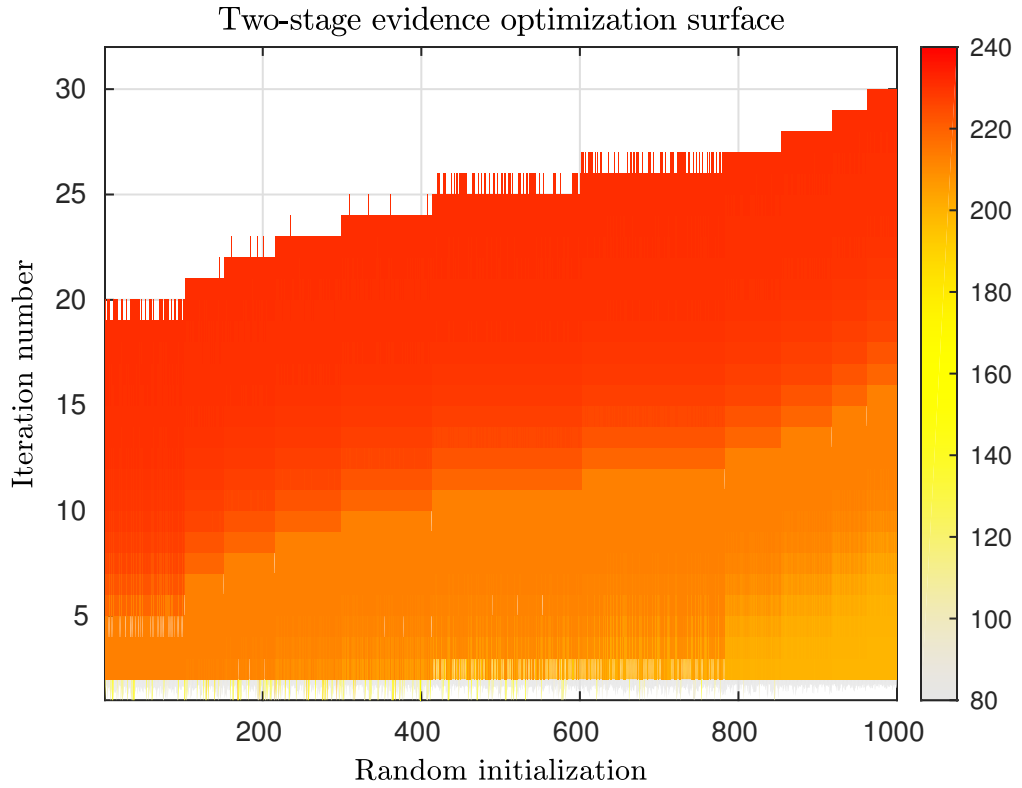


Figure 3.3. Two-stage evidence optimization surface as a function of the initial conditions (top-down view). The color denotes the log evidence. All the optimizations used the same synthetic data and an SNR of 4 dB. The simulations (bottom axis) were sorted by the number of iterations required for stage one of the algorithm to converge.

It is possible that although the algorithm converges to the same value of the log evidence, the resulting models could have different sparsity profiles (e.g., reflecting convergence to different local minima of the objective function). To assess this, in Fig. 3.4

we plot the hyperparameter estimates $\hat{\lambda}$ and $\hat{\gamma}_i$, following convergence, for each simulation sorted by the value of $\hat{\lambda}$. We augment these results by including the hyperparameter estimates yielded by the least squares initialization, whose location is denoted by the red circle in the bottom panel. A view to the bottom panel makes clear that λ converges to two different values. To measure the similarity of the sparsity profiles we compute one minus the cosine distance between pairs of $\hat{\gamma}$ vectors across simulation (Fig. 3.5). The top panel of Fig. 3.5 reveals that the algorithm converged to two slightly distinctive sparsity profiles. The bottom panel shows the source maps yielded by the hyperparameter set of three exemplary simulations. The left (initialized at random) and center (initialized by LS) panels show that the four simulated sources are recovered while the right panel shows an oversparsified map where the source in the middle frontal gyrus (R) is severely underestimated. We note that, although our algorithm is sensible to getting trapped in a local minima, the use of the LS initialization seems to place it on the path towards a plausible solution.

In Fig. 3.6 we compare the shortest (dash blue) and largest (dash-dot orange) optimization paths yielded by the random initial conditions shown in Fig. 3.3 and the path yielded by the LS initialization (solid black) using Eq. (3.13). The figure shows that the LS initialization produces an optimization path that is just a few iterations longer than the shortest path (within the models considered). Compared to the largest path, we see that stage one converges in half of the iterations, thereby contributing to speed up the algorithm. This experiment suggests that using Eq. (3.13) to initialize the hyperparameters of our model can remove delays due to suboptimal initial conditions.

3.3.4 Benchmark of the computational performance

BSBL methods have been shown to outperform many popular sparse solvers [Zhang and Rao, 2013]. In this experiment, we investigate the computational performance of the BSBL-2S algorithm in comparison with two closely related state of the art algorithms

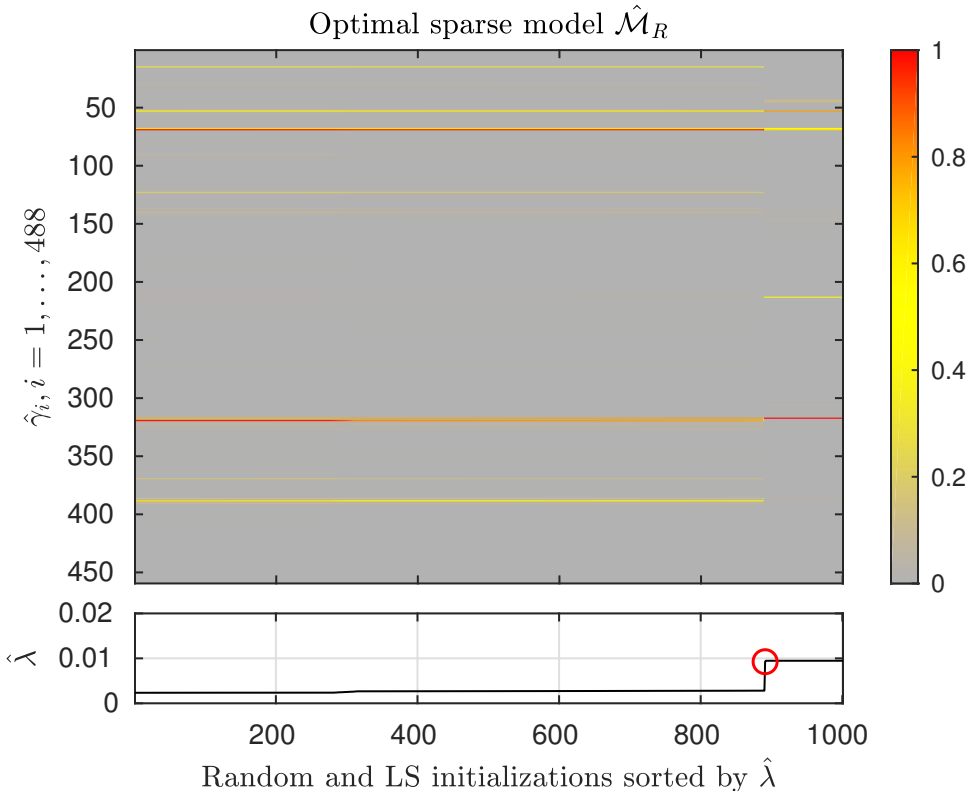


Figure 3.4. Values of the hyperparameters after algorithm convergence, sorted by the value of $\hat{\lambda}$ (bottom panel). In the top panel, the color of each pixel reflects the value of each $\hat{\gamma}_i$ parameter (y-axis) for the respective initialization (x-axis). The red circle in the bottom panel indicates the simulation initialized by least squares.

BSBL-EM and BSBL-BO [Zhang and Rao, 2013]. BSBL-EM uses update rules based on the EM algorithm and BSBL-BO uses the same λ update as the former and a γ update based on the optimization of an upper bound to the ML-II cost function that is essentially equivalent to γ-MAP. The benchmark simulates the online processing of EEG data blocks 80 milliseconds long. All algorithms were implemented fairly and optimized to reflect their minimum asymptotic complexity. Each benchmark used the same simulated ground truth data of Section 3.3.1 adding Gaussian noise accordingly.

In Fig. 3.7, the columns show benchmarks under different SNR from 2 dB to 10 dB. In each row, we plot the log evidence as a function of algorithm iterations (top) and milliseconds elapsed (bottom) respectively. The black dash line denotes the time point

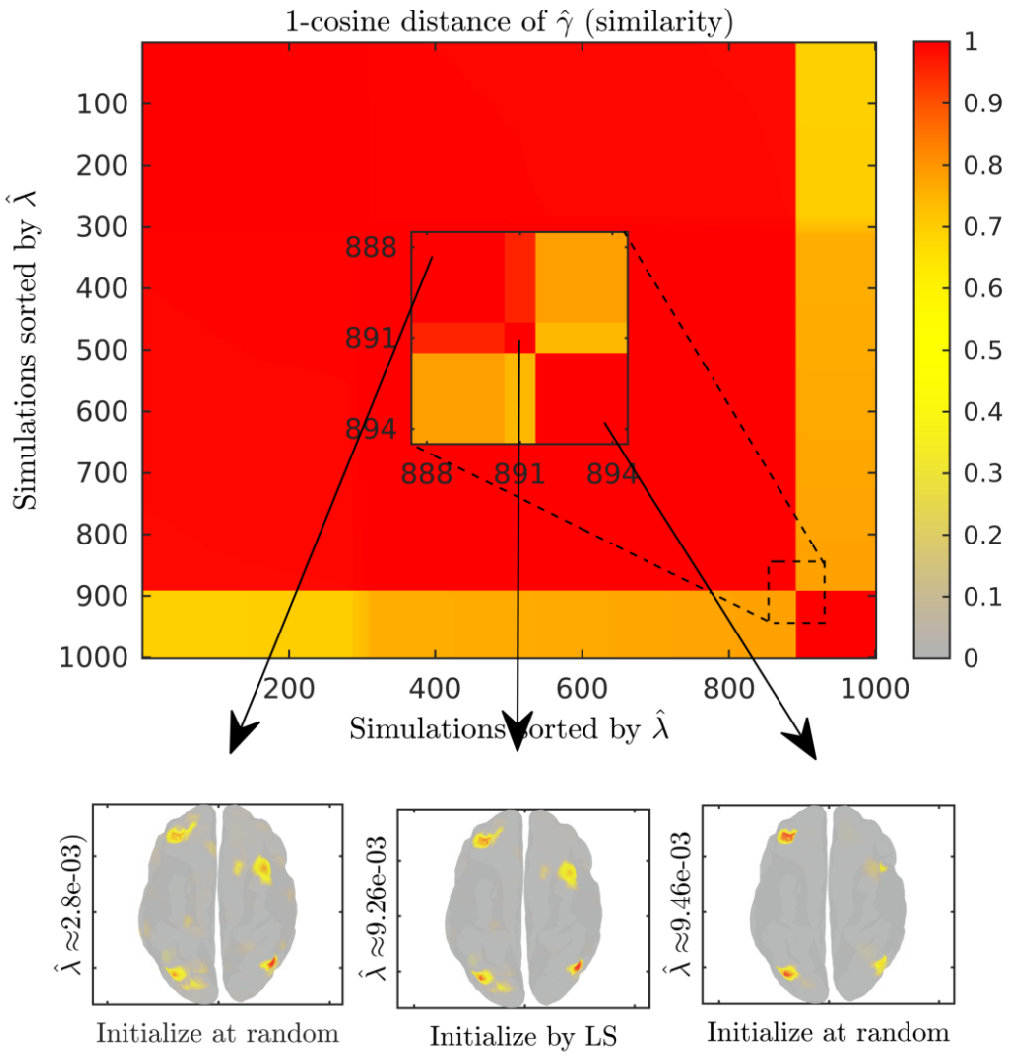


Figure 3.5. The top panel shows the similarity of every pair of $\hat{\gamma}$ vectors in Fig. 3.4. The similarity is measured as one minus the cosine distance between two $\hat{\gamma}$ vectors. The bottom panel shows the semitransparent source maps yielded by three exemplary simulations.

by which the subsequent batch of samples is available for processing. For an algorithm to perform in "real-time" it must converge before this mark. For the time span shown, from left to right we see that the log evidence increases as the SNR increased, intuitively indicating that the model uncertainty is reduced as the measurements are less affected by noise.

Inspecting the top row, we note that in some cases BSBL-2S requires only slightly

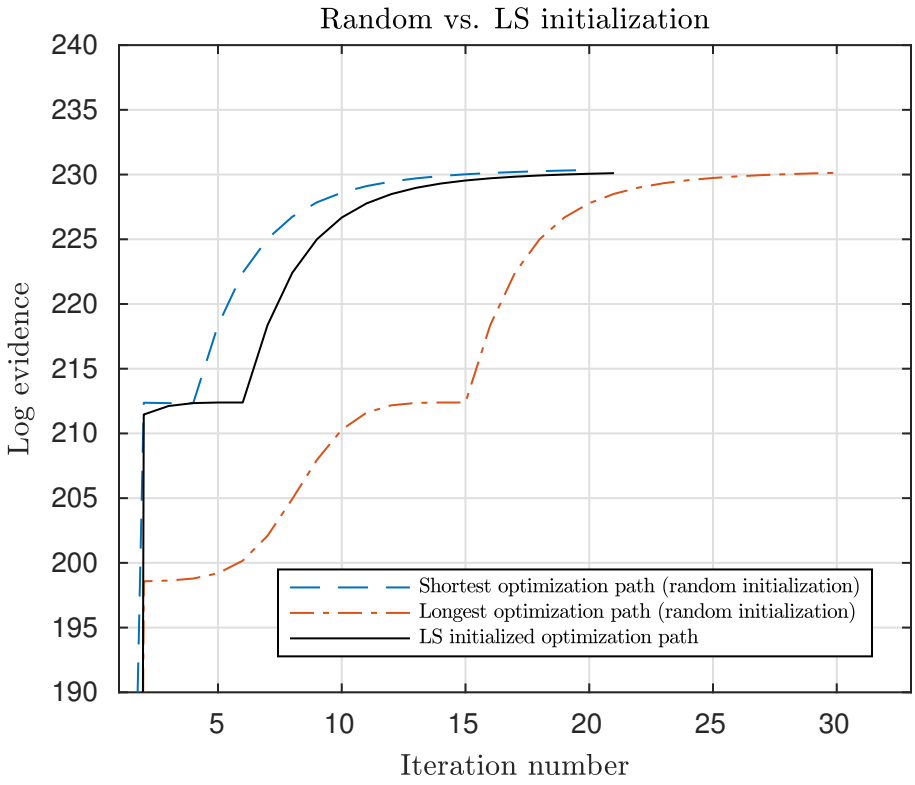


Figure 3.6. Detailed view of Fig. 3.3 showing the shortest and largest optimization paths yielded by random initial conditions and the one yielded by the LS initialization.

less iterations to converge than BSBL-BO. However, the bottom row panels demonstrate significant differences in the convergence time, with BSBL-2S significantly out-performing its counterparts. This suggests that the computational cost per iteration differs between algorithms. As indicated by the blue trace, we note the steep growth in log evidence achieved by stage one in the first iteration after the LS initialization, thereby contributing to the speed boost of BSBL-2S. As expected, in all cases the EM-based algorithm BSBL-EM displayed a slower convergence. We note that, in many practical online applications, source inference is one component of a pipeline wherein additional delays may be incurred by other signal processing or machine learning algorithms. As such, it is desirable for an algorithm to reach convergence in substantially less time than the minimum required for real-time processing.

Most SBL algorithms, including BSBL-BO and BSBL-EM, learn the hyperparame-

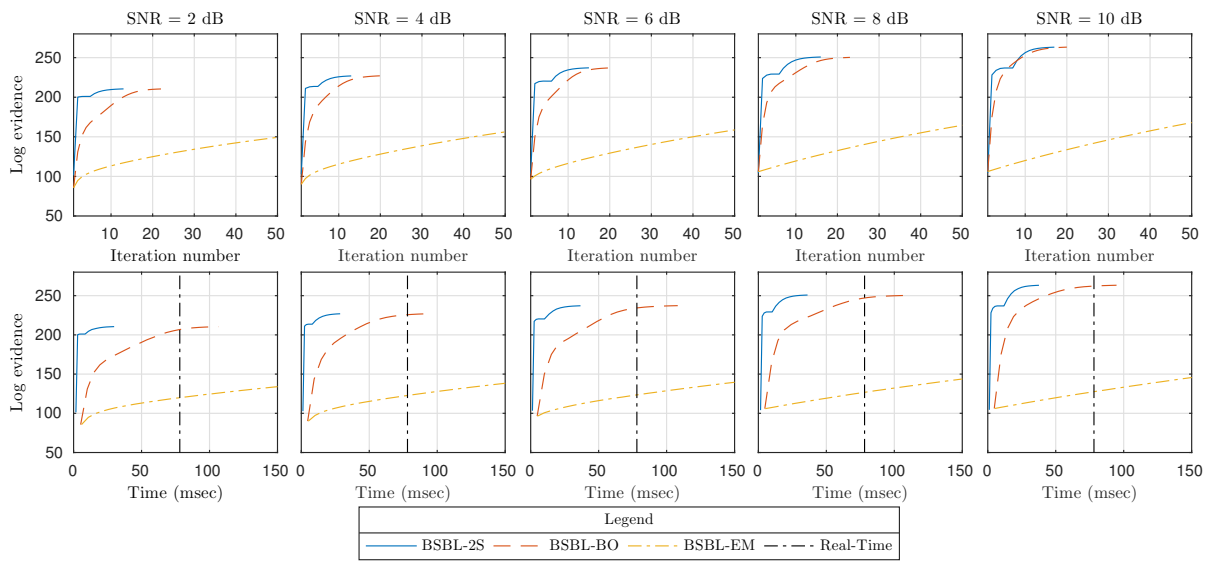


Figure 3.7. Benchmark of different SBL methods. All the tests used the same simulated data consisting of multiple sources concentrated within 4 ROIs scattered over the cortex. Each panel shows the benchmark under different SNR. The dash black line marks the moment when a new batch of samples is ready to be processed; to perform in real-time the algorithms must converge before this mark.

ters λ and γ simultaneously. BSBL-2S on the other hand, learns only two hyperparameters (λ, γ_F) in stage one, which from Fig. 3.7 seems to be quite efficient. In stage two, BSBL-2S learns the values of all N_g entries of the hyperparameter vector γ. We note that the fact that, in this simulation, BSBL-2S and BSBL-BO achieve approximately the same value of log evidence should be considered a happy accident and may not hold in general because in our algorithm the reduced model is never optimized with respect to λ. Nevertheless, our experiments show that the initialization of the second stage with the optimal full model has a positive impact on the convergence speed of the algorithm. Intuitively, we can see this process as an initial fast and reasonable, albeit coarse-grained, estimation, followed by a fine tuning step.

3.3.5 Adaptation to sensor noise fluctuations

In this example, we study the sensor noise adaptation characteristics of BSBL-2S. We simulated 1200 EEG samples, where after every 100 samples we changed the SNR by

varying the amplitude of the measurement noise, progressively changing the SNR from 10 dB to 0.5 dB and back to 10 dB. Sources were estimated in consecutive, contiguous, and nonoverlapping blocks of 10 samples. In Fig. 3.8, the blue trace represents the estimated noise precision (left y-axis) while the orange trace represents the SNR (right y-axis). The estimated noise precision is the inverse of the estimated noise variance $\hat{\lambda}$. We note that the estimated noise precision very closely tracked the changes in SNR, demonstrating that BSBL-2S can adapt to changes in sensor noise.

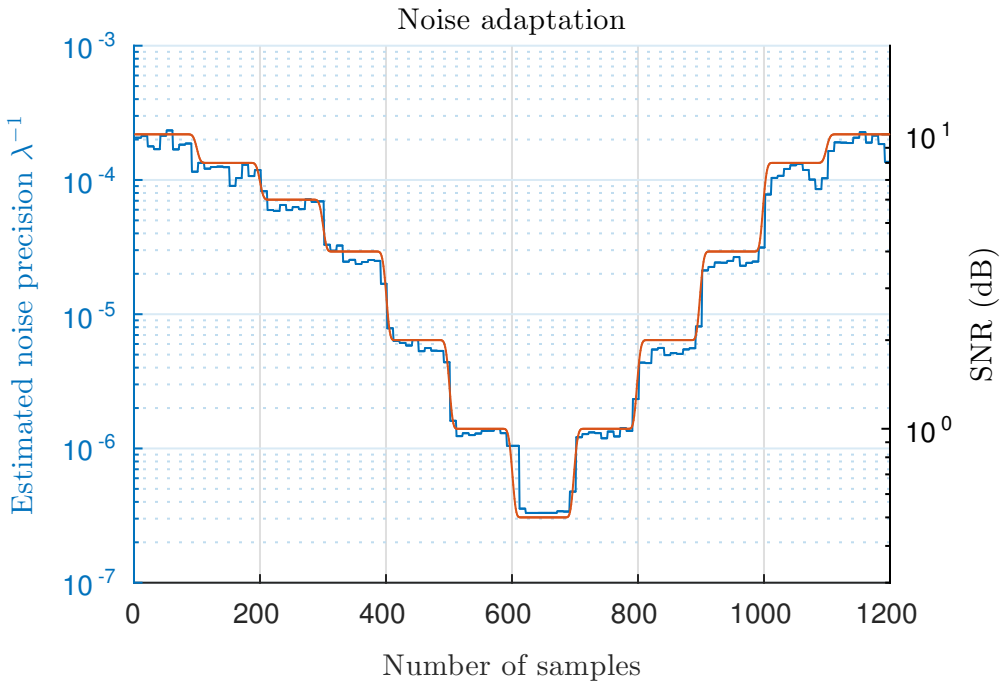


Figure 3.8. The blue trace represents the inverse of the hyperparameter $\hat{\lambda}$ estimated every 10 consecutive samples. The orange trace represents the simulated changes in SNR.

3.3.6 Benchmark of the source reconstruction quality

With the objective of reducing computational cost during real-time source estimation, some authors have proposed fixing the regularization parameters to values obtained offline [Pieloth et al., 2014]. Following up on the results of Section 3.3.5, in this example we show that the lack of features such as online adaptation and sparsity can significantly reduce the quality of the recovered sources. We compared the BSBL-2S algorithm against

the following inverse solvers that can be readily obtained within the framework outlined in this chapter:

- BSBL-2S (λ fixed): Same as BSBL-2S but λ is fixed after the first 10 samples. In this case the algorithm does not adapt to changes in sensor noise.
- LORETA: optimization of the full model only (stage one) and the pruning stage (stage two) disabled. This method estimates sources subject only to smoothness constraints but adapting λ and γ_F.
- LORETA (λ fixed): same as LORETA but with λ updates turned off after the first 10 samples.

We do not include BSBL-BO and BSBL-EM algorithms in the comparison because, as shown in Fig. 3.7, they do not perform in real-time. In particular, we note that for a fixed λ, BSBL-2S and BSBL-BO algorithms reduced to the same procedure (γ-MAP).

In Fig. 3.9, the top panel shows the performance achieved by each inverse method. The bottom panel indicates groups of samples with different SNR, which for this simulation followed the same pattern as in Fig. 3.8. In all cases, SBL methods outperformed LORETA. This is not surprising given the tendency of LORETA to overestimate the spatial extent of the activations and produce “ghost" (spurious) source estimates. For samples with high SNR, we can see the performance boost introduced by the pruning step, highlighting the importance of using sparsity constraints. The simulation shows that BSBL-2S was the only method robust enough to extreme noise conditions (SNR $\leq$ 1 dB), demonstrating the importance of not only imposing sparsity constraints (γ inference) but also adapting to changes in sensor noise (λ inference) in a data-driven manner.

3.3.7 Source estimation of error-related potentials

In this section, we use the BSBL-2S algorithm for computing single-trial estimates of the sources underlying error-related EEG activity from a healthy subject performing

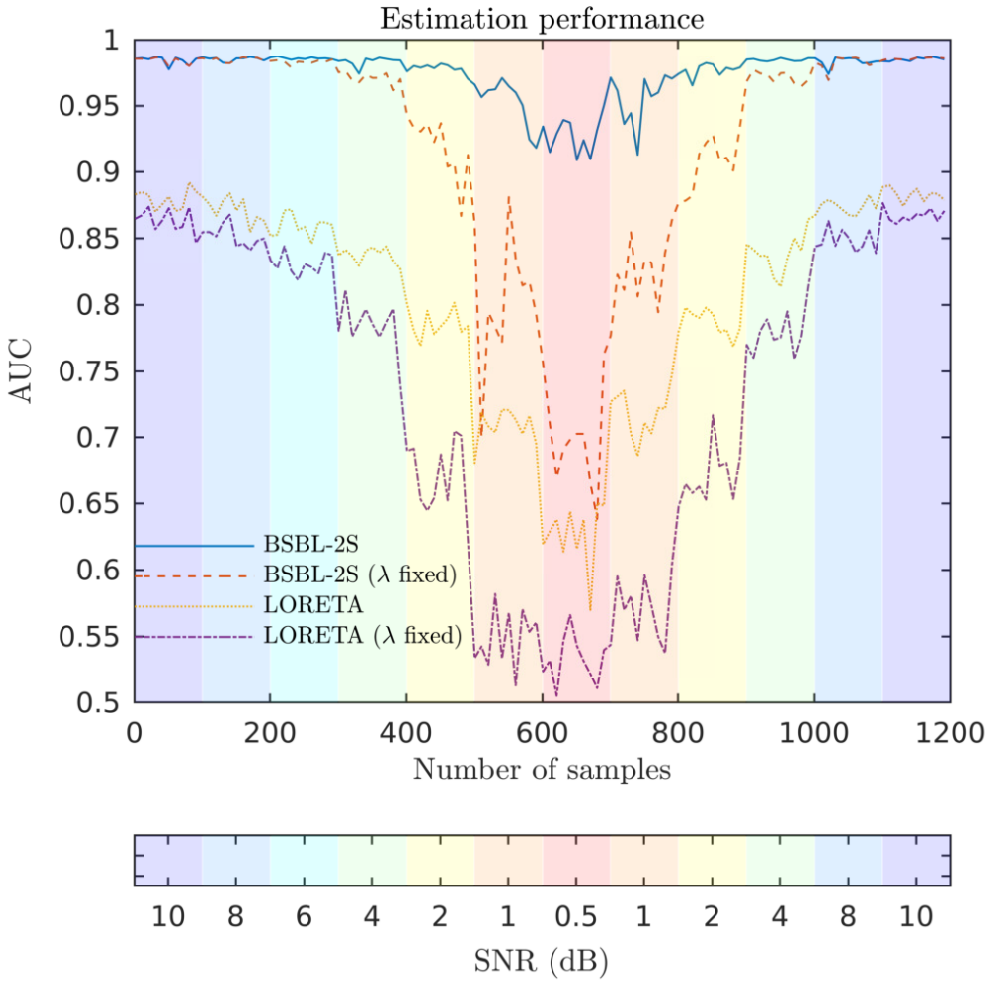


Figure 3.9. The traces show the AUC of different inverse methods. The colored bars represent groups of samples with different SNR, from 10 dB to 0.5 dB.

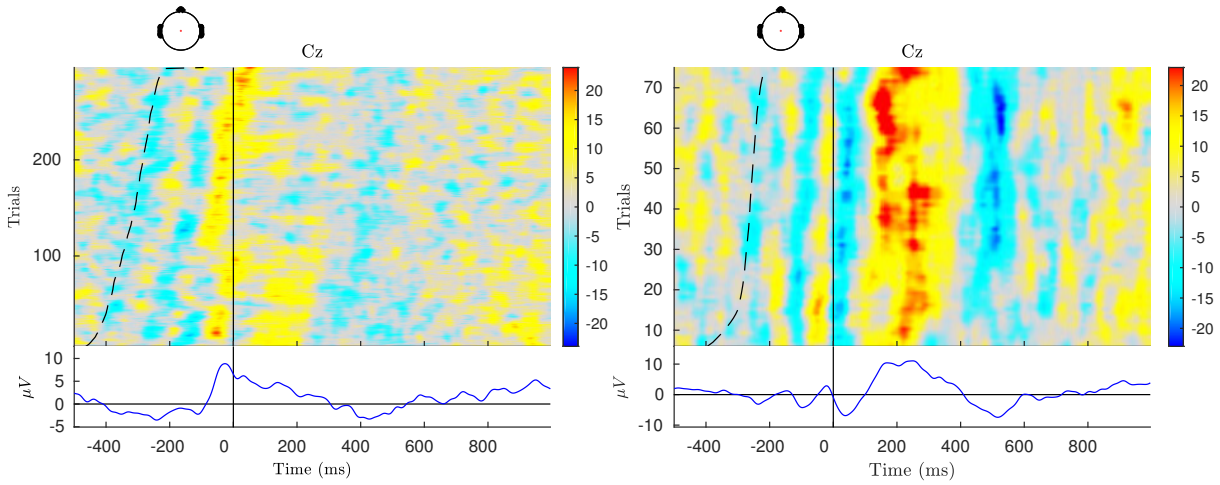


Figure 3.10. Single-trial EEG activity correspondent to the correct (left panel) and error (right panel) conditions and averaged ERP for sensor Cz. Response-locked trials are sorted by the time elapsed from the presentation of the target (dashed trace) to the onset of the response (solid trace). The cartoon head on the top left shows the location of sensor Cz.

a modified Flanker task. The data were kindly provided by McLoughlin et al. [2014]. Data were recorded using a 64-channel (BrainAmp DC; Brain Products, GmbH, Munich, Germany) extended 10-20 system montage (reference at FCz) digitized at a 500 Hz sampling rate. The subject was presented with visual stimuli consisting of two flankers (horizontal arrows pointing either to the left or the right) above and below a fixation mark. 100 milliseconds later a central target arrow appeared for a duration of 150 milliseconds. The subject was instructed to press a response button with the index finger of the hand (left or right) corresponding to the direction indicated by the target arrow (left or right). On congruent trials, the flanker and target arrows pointed in the same direction, while on incongruent trials they pointed in opposite directions. 10 blocks of 40 trials were recorded. For further details on the task and data recording, we direct the reader to [McLoughlin et al., 2014].

In this paradigm, after the commission of an error, a response-locked error-related negativity (ERN) ERP is often elicited, also referred to as Ne. Typically, the Ne shows a negative peak in the frontocentral sensors within 100 milliseconds after the erroneous

response. The Ne component is often followed by a pronounced positivity in the centroparietal sensors around 200-500 milliseconds, which is referred to as Pe. We test our method on this type of data because error-related potentials have been studied in great detail [Gehring et al., 2012] and they have practical relevance for BCIs [Chavarriaga et al., 2014].

We carried out standard data pre-processing within the EEGLAB software environment [Delorme et al., 2011]. We removed electrooculogram (EOG) channels yielding 60 EEG channels. We then applied a 0.5 Hz to 35 Hz band-pass forward-backward FIR filter, downsampled to 250 Hz, and re-referenced the data to the common average. We removed eye-blink artifacts using the online-capable Artifact Subspace Reconstruction method [Mullen et al., 2015]. EEG data were epoched into trials spanning -500 to 1000 milliseconds relative to the onset of correct and erroneous responses and the mean baseline over -500 to 0 milliseconds was subtracted. The erroneous trials included those were the stimulus were both congruent and incongruent. The final data set consisted of two tensors $\mathcal{Y}_{c,e} \in \mathbb{R}^{N_y \times N_s \times Tc,e}$, where $N_y = 60$ denotes the number of sensors used for analysis, $N_s = 250$ is the number of samples in each trial, and T is the number of trials ($T_c = 305$ correct and $T_e = 74$ erroneous).

To construct a head model, we first nonlinearly warped the subject's sensor locations to the scalp surface of the template using the *headModel* toolbox[3]. In contrast to previous experiments on simulated data, in this section we computed the orientation free lead field, i.e., we estimated (x, y, z) components for each source for a total of $3N_g = 15009$ parameters. The group sparsity constraints were imposed along every (x, y, z) direction independently, i.e., sources were allowed to have a trivial component along one or more directions. We estimated the source activity along the time dimension of the $\mathcal{Y}_{c,e}$ tensors independently for every trial. We applied BSBL-2S to batches of 80 milliseconds (20 samples) data with 50% overlapping to smooth out high-frequency components introduced by sharp transitions in the sparsity profile of the sources from one block to the next.

[3]https://github.com/aojeda/headModel

Fig. 3.10 shows correct (left panel) and erroneous (right panel) single-trial EEG responses for sensor Cz as ERP-images[4] [Delorme et al., 2015]. In each panel, we sorted trials by the time elapsed from the presentation of the target arrow (dashed trace) to the onset of the response (solid trace). The bottom panel shows the ERP obtained by averaging the trials in the top panel. As shown in the panel on the right, we obtained an Ne component with maximum negativity at 68 milliseconds, followed by the Pe component with a peak at 252 milliseconds.

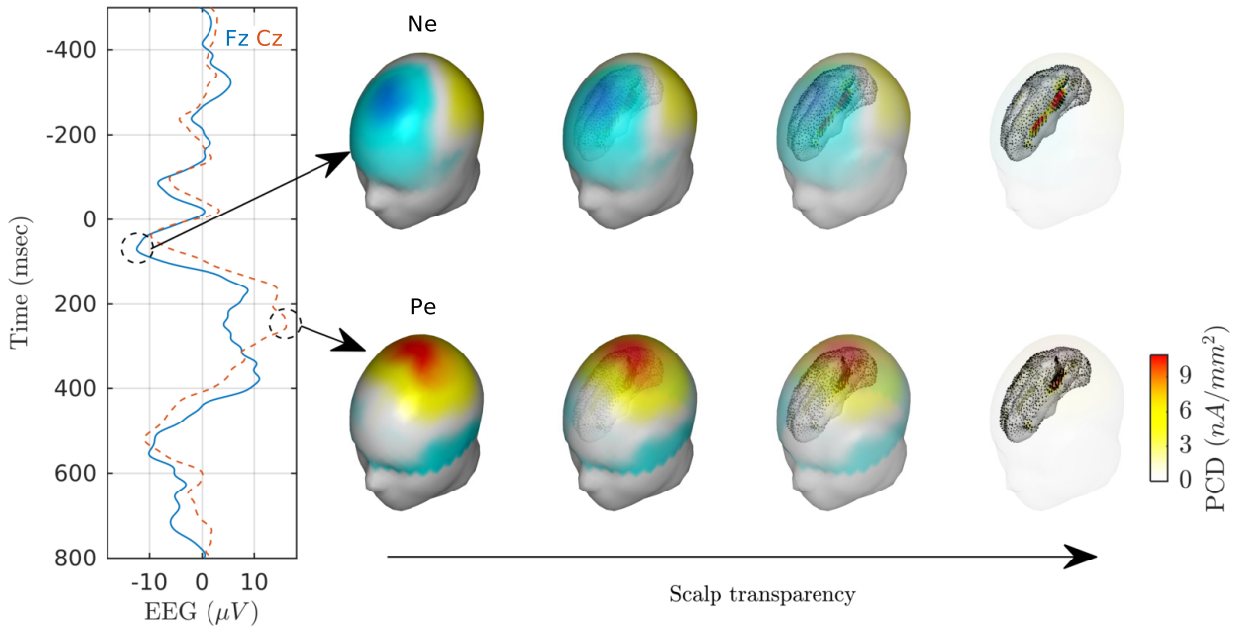
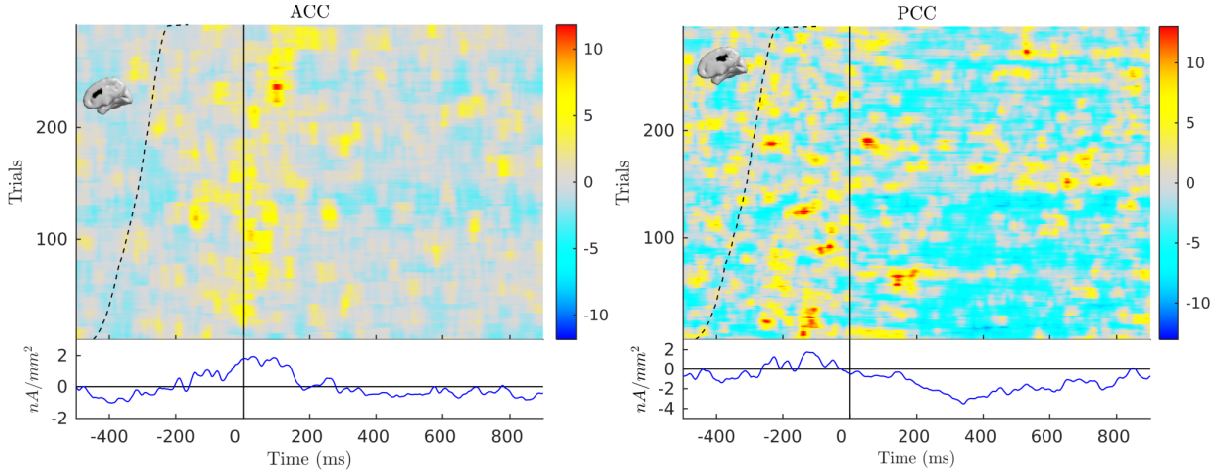


Figure 3.11. Source estimates averaged over error trials. The leftmost panel shows the ERP time courses for sensors Fz and Cz. The dashed circles mark the latency of Ne/Pe components and each arrow points to the respective mean scalp topography (interpolated on the surface of the head) and their underlying cortical PCD maps. The bottom arrow indicates that the scalp transparency fades out from left to right, allowing us to see in the center panels the superposition of scalp topographies and cortical activations. The orientations of the source dipoles are indicated by black arrows on the cortical surface. A movie detailing the scalp and source ERP activity can be found https://www.dropbox.com/s/nuqwkviy4d9khfo/movie_Fig11.avi?dl=0.

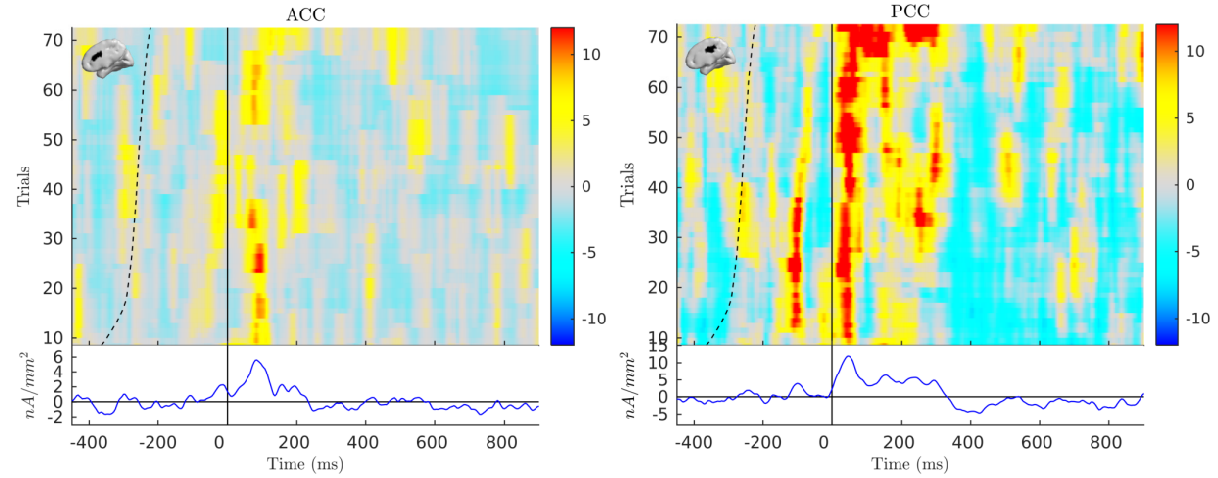
Averaging the single-trial source images for the error condition obtained by BSBL-2S

[4]In this section, the lower and upper colormap limits reflect the 5th and 95th percentile, respectively, over the displayed images.

revealed consistent activity within two brain regions that have been linked to conflict monitoring and error processing, the anterior cingulate cortex (ACC), and the posterior cingulate cortex (PCC) [Buzzell et al., 2017; Herrmann et al., 2004; Luu et al., 2004; Roger et al., 2010; Vocat et al., 2008]. We found that the generators of the Ne component included sources located in both ACC and PCC regions pointing away from the sensors, while the Pe component was generated by sources in the PCC pointing towards the sensors (see Fig. 3.11).



(a) Correct trials.



(b) Erroneous trials.

Figure 3.12. Baseline corrected single-trial source estimates magnitude averaged within the ACC and PCC regions.

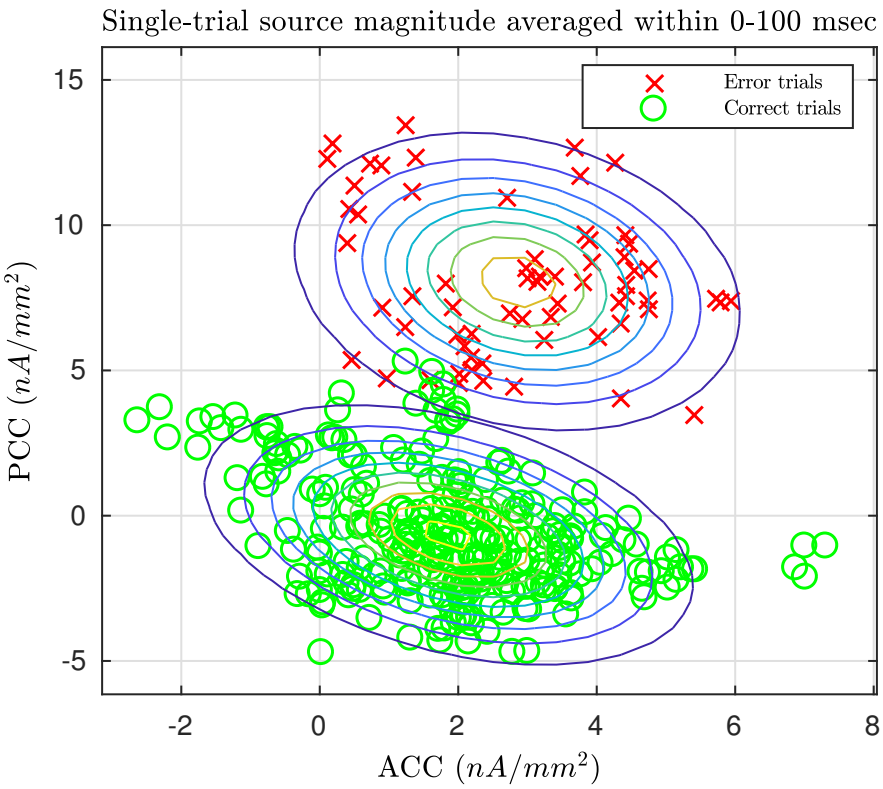


Figure 3.13. Two Gaussian mixture model fit of baseline-corrected source magnitudes from ACC and PCC in the 0-100 msec interval for error (red ×'s) and correct (green circles) trials. The colored contour lines represent regions of equal probability.

Finally, we present a simple example illustrating the use of brain-localized single-trial source estimates for identifying error-related brain responses. Fig. 3.12 plots the baseline-corrected magnitude of the single-trial source estimates averaged within the ACC and PCC regions for the correct (a) and error (b) conditions. We further averaged the source magnitude within the first 100 milliseconds after the response and plotted each trial as a data point in an (ACC, PCC) plane (see Fig. 3.13). In Fig. 3.13, the green circles and red × symbols represent correct and error trials respectively and the colored contour lines represent regions of equal probability obtained by fitting a two Gaussian mixture model to the data. Although the issues pertaining the learning of a classifier are beyond the scope of this chapter, the proposed method may be useful for obtaining features with increased biological relevance compared to sensor-based features.

3.4 Conclusions

In this chapter, we have developed BSBL-2S, a new Sparse Bayesian Learning (SBL) algorithm that can deliver fast, block-sparse, and robust solutions to the EEG source imaging (ESI) problem. Our contribution can be seen as a sensor noise adaptive pre-training for the γ-MAP algorithm of Wipf and Nagarajan [2008] that allows for real-time batch processing of EEG samples. To achieve this in a computationally efficient manner, we decouple the estimation of the sensor noise covariance and the sparsity profile of the sources, yielding a fast two-stage algorithm. In the first stage, we optimize a simplified non-sparse generative model to get an estimate of the sensor noise covariance and a good initialization of sources group-sparsity pattern. Sources obtained at this stage are equivalent to those estimated with the inverse method LORETA. In the second stage, we apply the γ-MAP algorithm with the noise covariance fixed to the value obtained in stage one to efficiently shrink to zero groups of sources that are irrelevant for explaining the observed EEG data. Furthermore, we derive an initialization to the stage one of the algorithm that is optimal in the least squares sense, thereby preventing delays due to suboptimal random initialization parameters.

On 64-channel simulated EEG data, we showed that the BSBL-2S algorithm can perform in real-time and that the two-stage approach improved the convergence speed relative to comparable single-stage SBL algorithms. We additionally demonstrated that the algorithm can robustly adapt to fluctuations in the sensor noise variance. On real EEG data, we localized the Ne and Pe components of single-trial error-related potentials to compact regions of the cortex that are in agreement with previous neuroimaging studies. We believe that the proposed approach will prove useful for applications within and outside of the neuroimaging and brain-computer interface fields, thereby contributing to the use of EEG as a brain imaging modality.

3.5 Acknowledgment

The authors wish to gratefully acknowledge our collaborator Jean Vettel for her helpful comments. We thank the authors of [McLoughlin et al., 2014] for providing the real data used to validate our method. We also thank the anonymous reviewers for their thoughtful comments.

This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0022. The views and the conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S Government. The U.S Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

In part, this chapter is a reprint of the material as it appears in Ojeda, A., Kreutz-Delgado, K., & Mullen, T. (2018). Fast and robust Block-Sparse Bayesian learning for EEG source imaging. NeuroImage, 174(8), 449–462.
https://doi.org/10.1016/j.neuroimage.2018.03.048. I was the primary investigator and author of this paper.

Chapter 4

SimBSI: An open-source Simulink library for fast development of cross-species closed-loop brain signal interfaces

4.1 Introduction

We can define a brain-computer interface (BCI) as a bidirectional communication channel between a brain and a computer with the purpose of compensating dysfunctional neuronal activity [Liang et al., 2010], rehabilitating motor skills [Hurtier et al., 2016], or enhancing a cognitive ability [Burke et al., 2015]. Likewise, a closed-loop BCI can be defined as a BCI that uses stimulation and feedback loops to adapt its communication channel to ongoing changes in brain dynamics, task goals, and environmental signals. A promising application of closed-loop BCI technology to clinical neuroscience is the development of personalized computerized therapies that can target specific neural circuit dysfunctions leading to psychiatric disorders [Carelli et al., 2017; McFarland et al., 2017; Mishra and Gazzaley, 2014]. Most common BCI applications, however, offer limited therapeutic value due to oversimplified designs and poor understanding of the neurobiological basis of the mental health conditions being treated [Jeunet et al., 2016; Lotte et al., 2013].

Building BCI therapies on more solid grounds may require the characterization of

the brain dynamics supporting cognition and behavior at multiple scales, from single-cell and local field potential (LFP) recordings in animals to non-invasive electroencephalogram (EEG), functional near-infrared spectroscopy (fNIRS), and functional magnetic resonance imaging (fMRI) in humans [Mishra and Gazzaley, 2016]. For instance, taking advantage of complementary species-specific methods, cross-species studies have shed light on the neurophysiological mechanisms of fear and anxiety [Malter Cohen et al., 2013; Pattwell et al., 2012; Soliman et al., 2010], rapid learning [Sagi et al., 2012], and error-related adaptive learning [Narayanan et al., 2013]. Furthermore, BCI training has shown to diminish distractibility in aging rats and humans [Mishra et al., 2014].

Despite recent efforts [Siegle et al., 2017], software and hardware supporting cross-species closed-loop studies are still lacking. In this chapter, we introduce Simulink for Brain Signal Interfaces (SimBSI), an open-source library and auxiliary programs for facilitating closed-loop BCI research in rodents and humans. Simulink is a graphical programming environment for modeling and simulation of dynamical systems and signal processing that is tightly integrated with MATLAB (The MathWorks, Inc., Natick, Massachusetts, USA), which has gained popularity in recent years for implementing EEG [Chiesi et al., 2018; Georgieva et al., 2014; Guger et al., 2001] and single-cell [Zrenner et al., 2010] BCIs. SimBSI adds to this ecosystem advanced human brain mapping methods for monitoring and targeting cortical regions in real-time based on EEG signals, flexible cross-species multimodal data acquisition based on the Lab Streaming Layer (LSL) library [UCSD Swartz Center for Computational Neuroscience, 2011], as well as a flexible experimental design platform for developing closed-loop BCI systems.

The rest of the chapter is organized as follows. In Section 4.2 we explain the design principles that we followed to deveop the SimBSI library. In Section 4.3 we explain briefly how Simulink works in the context of real-time signal processing applications. In section 4.4 we give an overview of multimodal, multirate data acquisition and synchronization using the LSL library. Sections from 4.5-4.7, comprise a series of examples for signal processing,

Table 4.1. BCI environments reviewed for this chapter compared against SimBSI.

Software	Language	Graphical prog.	Built-in LSL	Android and iOS	Raspberry Pi	Arduino	Ref.
BCI2000	C++	✗	✗	✗	✗	✗	[Schalk et al., 2004]
AsTeRICS	Java, C/C++	✓	✗	✗	✗	✗	[Drajsajtl et al., 2013]
OpenViBE	C++	✓	✓	✗	✗	✗	[Renard et al., 2010]
BCILAB	MATLAB	✗	✓	✗	✗	✗	[Kothe and Makeig, 2013]
Pyff	Python	✗	✗	✗	✓*	✗	[Venthur et al., 2010]
TOBI	C++	✗	✗	✗	✗	✗	[Müller-Putz, 2011]
BCI++	C/C++, C#, MATLAB	✗	✗	✗	✗	✗	[Perego et al., 2009]
xBCI	C++	✓	✗	✗	✗	✗	[Susila et al., 2010]
BF++	C++	✗	✗	✗	✗	✗	[Quitadamo et al., 2008]
gumpy	Python	✗	✓	✗	✓*	✗	[Tayeb et al., 2018]
SimBSI	MATLAB, C/C++	✓	✓	✓	✓	✓	This chapter

*Although it is not specified by the authors, a simplified version of this software may run on a Raspberry Pi board.

EEG source imaging, cross-species cognitive task design, and closed-loop neuromodulation. Finally, in Section 4.9 we describe a hardware/software environment for closed-loop animal BCIs.

4.2 Design principles

Over the last two decades, we have witnessed a rapid development of miniaturized, high-precision, and low-cost bio-sensing technologies [Mcdowell et al., 2013] accompanied by hardware of increased computing power. These advancements have fulled the creation of several software environments dedicated to human BCI where practitioners can implement a variety of predefined as well as customized approaches [Brunner et al., 2012]. Ideally, one would like to extend existent human BCI systems, such as those shown in Table 4.1, to support cross-species experiments. Extending existent BCIs to other species, however, may be a cumbersome process and in many cases unfeasible due to differences in hardware, data modalities, experimental paradigms, and programming languages.

SimBSI is not a BCI software per se, i.e., it does not offer popular predefined approaches, nor is it optimized with a specific data modality in mind. Instead, we designed it as a library that implements functionality for biosignal acquisition and processing that complements the broad set of tools already available in Simulink for online signal processing (see Fig. 4.1). In particular, we designed SimBSI with the following principles in mind:

- Intuitive programming environment: many BCI practitioners come from fields where

mastering a programming language is not a requirement. So, not having to code in the traditional way, but graphically, automatically lowers the technical background required to design a BCI.

- Transparency of data processing: the flowchart nature of the Simulink language makes it straightforward to debug and document a processing pipeline.

- Multiplatform: most Simulink programs can be compiled into standalone apps and deployed to 1) traditional platforms such as Windows, Mac and Linux, 2) embedded hardware such as Arduino and Raspberry Pi, 3) mobile devices such as Android, iPhone, and iPad.

- Flexible data acquisition: Simulink's Instrument Control Toolbox allows interfacing instruments using standard communication protocols such as TCP/IP, UDP, and the serial port.

- Reuse as much code as possible: large frameworks require much effort to develop, maintain, document, and validate. Since Simulink offers a well validated general-purpose online signal processing environment, why start from scratch? Furthermore, Simulink can be extended with new blocks using C/C++ or standard MATLAB code, so it is possible to reuse a large set of neuroscientific tools already developed and tested in MATLAB. For example, popular human BCI approaches could be ported from BCILAB [Kothe and Makeig, 2013] or BioSig [Schlögl and Brunner, 2008] toolboxes.

4.3 Simulink in a nutshell

In this section we outline the fundamentals of how Simulink works. This is important because it will help us to understand the capabilities and limitations of Simulink as a platform for closed-loop BCI and how to circumvent them. We base our exposition upon

what we have learned from MathWorks' online documentation and we refer the reader to it for more in-depth explanations.

To write a program in Simulink, we use its graphical editor to connect different blocks to form a flow diagram (or pipeline) that represents the time-dependent operations performed to the signals that propagate through it. Once the pipeline is designed, we set the sample time at which the system will operate (Model Configuration Parameters → Solver → Fixed-step size); this is usually set as the inverse of the highest sampling rate of the signals acquired by the pipeline. We set also the stop time in the toolbar of the editor (also exposed in Model Configuration Parameters → Solver → Stop time); this is the total time that the system will run for. For testing, the stop time could be any number of seconds, and for deployment, we can set it to a large number, e.g., 3600 if we want it to run for 1 hour or *inf* if we want it to run indefinitely.

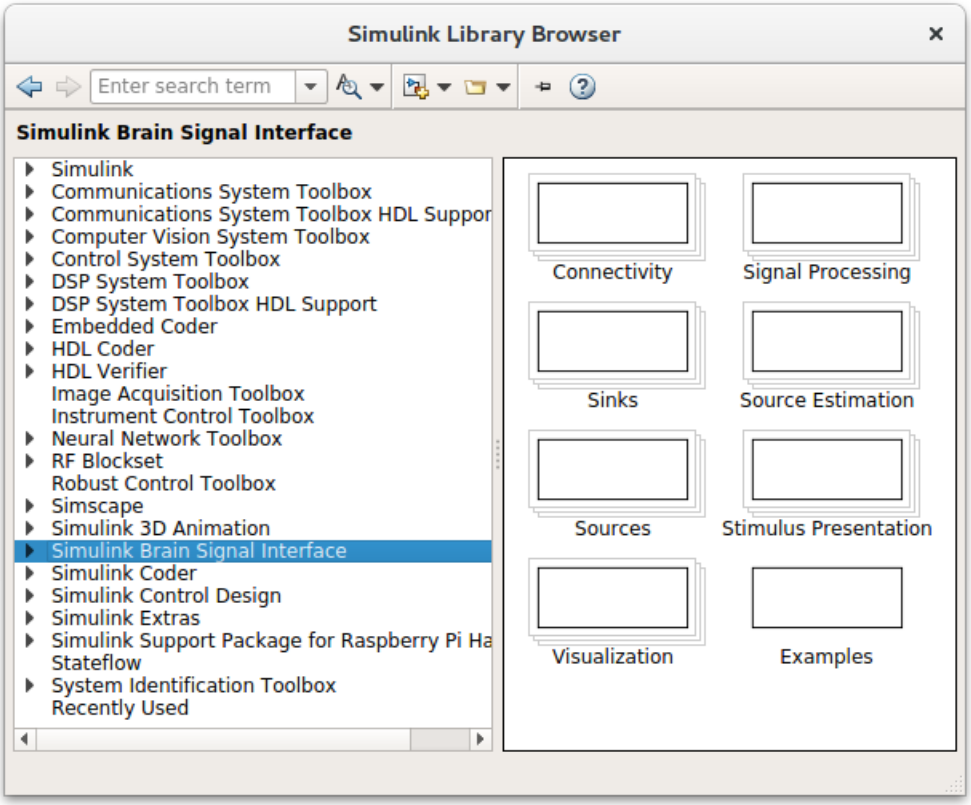


Figure 4.1. SimBSI shown inside the Simulink library browser.

In a closed-loop experiment, it is paramount to control the rate at which a pipeline is executed so that we can deliver feedback to the subject during the targeted neural state. To make sure that a pipeline runs in real time, during the prototyping phase of a project it is a good practice to use the performance tools in the editor (Analysis → Performance Tools), e.g., processing 10 seconds of data should take up to 10 seconds but no more. Depending on the complexity of the pipeline it may be possible to achieve real-time performance in the default mode of operation, which is called "normal mode". In normal mode, the pipeline can be executed combining compiled and interpreted blocks, thus facilitating rapid prototyping, testing, and debugging. For soft real-time deployment, however, we would typically use an accelerated mode [MathWorks, 2018a]. In accelerated modes, all blocks are translated into C code and the pipeline is compiled into a binary.

If additional performance is needed beyond what can be achieved in accelerated modes, there are two more options: the Simulink Desktop Real-Time (SDRT) Toolbox and Simulink Real-Time (SRT). SDRT technology works by turning a general-purpose computer into a real-time system during the execution of a Simulink pipeline. It does so by installing a real-time kernel that takes over the CPU at run-time so that I/O and other time-critical operations can be performed without sharing CPU time with unrelated processes. After completing each execution step, the kernel yilelds the CPU to serve the processes put on hold during pipeline execution. SRT, on the other hand, needs a host and a target computer. The host is used to design and tune the pipeline before it is deployed onto a dedicated (target) hardware, which can be a high-performance computer, FPGA, or DSP board for processing high volumes of data at shorter sample times. This technology also allows routing signals from the target back to the host computer for online visualization without compromising real-time performance.

4.4 Multimodal data acquisition with LSL

As we mentioned earlier, Simulink offers several blocks for standard instrument communication. For instance, in a closed-loop experiment, a pipeline could read in LFP/EEG samples via the TCP/IP protocol, filter them, extract neuromarkers, and send out a feedback stimulation signal if a target state is detected. If in addition, we need to save signals to disk or account for time delays between different sensors or stimulation devices, a common practice is to time stamp every sample using the clock of the LFP/EEG amplifier. To record experimental and behavioral events, it is a common practice to use TTL pulses connected to the auxiliary channels of the amplifier used to record the electrophysiological data.

In experiments involving the distributed collection of multiple data modalities at different rates, stimulation devices, and control signals, the synchronization via TTL pulses can become nontrivial. In these cases, a much simpler approach is to link data acquisition, stimulus presentation, and stimulation hardware with the LSL library. LSL is built on top of the TCP protocol and is designed to synchronize different devices that publish their data on the same local area network (LAN). Each sample received by the library gets assigned a time stamp calculated taking into account the offsets of the different clocks participating in the experimental session. Furthermore, the time synchronization mechanism implemented in LSL was designed after the widely used Network Time Protocol (NTP) with the objective of achieving sub-millisecond accuracy. Thanks to its extensive documentation and easy to use API, LSL has gained popularity within the EEG community and now is supported by several hardware makers [UCSD Swartz Center for Computational Neuroscience, 2018b].

To our knowledge, so far LSL is virtually unknown in the field of animal electrophysiology. We believe, however, that its use could greatly simplify complex animal experimental setups. Also, recording data using the same underlying technology and file formats would enable the use of common software tools for online and offline signal

processing and analysis in cross-species studies. For these reasons we have written an LSL plugin for the Open Ephys platform [UCSD Neural Engineering and Translation Labs, 2018a] (see Fig. 4.2). Open Ephys is a popular open-source plugin-based framework mostly used for multichannel animal electrophysiology [Siegle et al., 2017]. This plugin is not part of the SimBSI library but it is an essential component of the LSL ecosystem that enables the development of closed-loop animal experiments using Simulink and the tools in SimBSI.

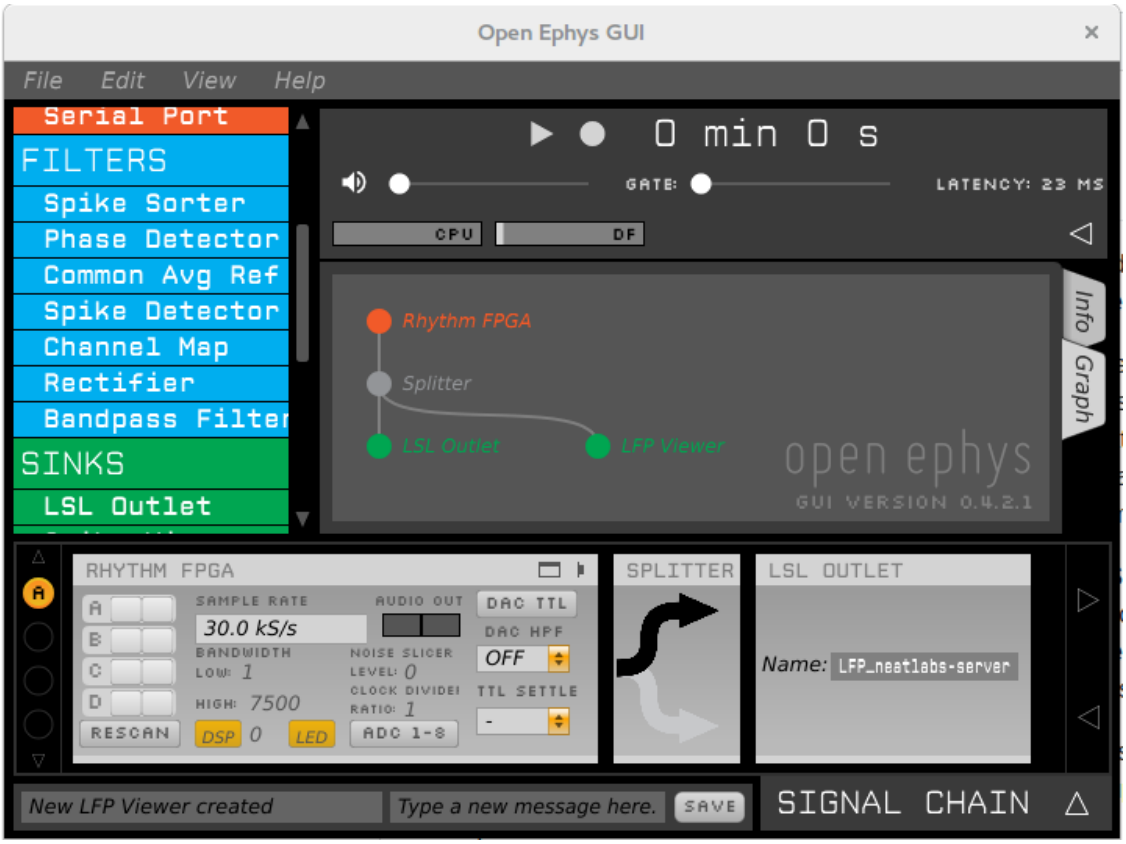


Figure 4.2. LSL plugin for Open Ephys GUI. The configuration shows Open Ephys connected to the RHD2000 FPGA Rhythm board (Intan Thechnologies, Inc., USA) for 30 kHz data acquisition from implanted electrodes and visualization. Every small batch of data acquired from the board is forwarded to LSL for online analysis or storage to disk.

To access LSL data streams within the Simulink environment we created the *LSLInlet* block. The *LSLInlet* block is a Simulink extension function (s-function) written in C using LSL and Simulink's APIs. Fig. 4.3 **A** shows an example of EEG data acquisition

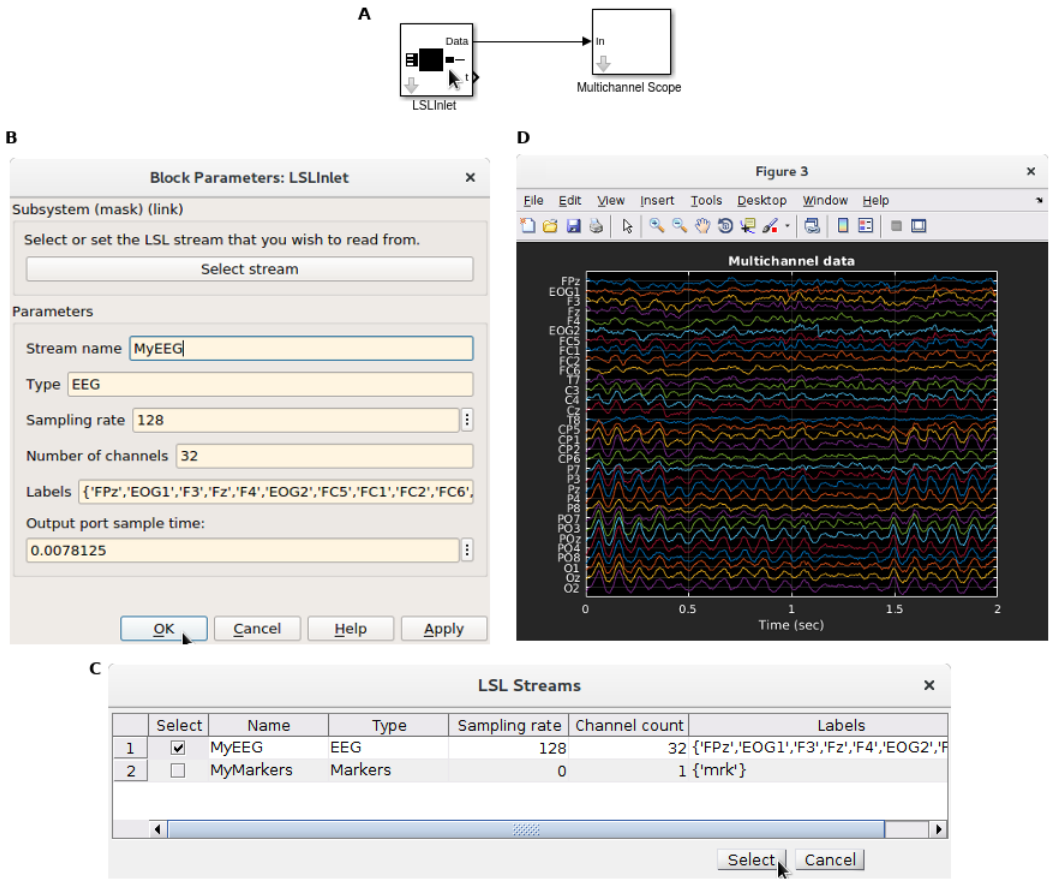


Figure 4.3. EEG data acquisition and visualization. **A:** Simulink pipeline that uses the *LSLInlet* block for reading data in from the network and the *Multichannel Scope* block for visualization. **B:** LSL configuration window. **C:** Stream selection tool that pops up when we click on *Select stream*.

using the LSLInlet block, which can be found in SimBSI → Sources. To configure the LSLInlet block we double click on it and set the stream parameters in the configuration window (panel **B**). For convenience, all the parameters can be filled up automatically by clicking on the *Select stream* button and selecting the desired stream from the list of all stream available in the LAN we are connected to (panel **C**). Panel **D** shows 2 second snapshot of EEG data flowing through the pipeline. After the sampling rate is set, i.e., to 128 Hz, the *LSLInlet* block automatically sets the sample time of the pipeline to 0.0078125 (1/128) sec. We note that in normal mode, Simulink typically runs as fast as it can regardless of the set sampke time, however, the use of the *LSLInlet* block forces Simulink

to wait until a new sample is available in LSL, thereby guaranteeing an effective execution time equal or higher, but not faster, than the sample time of the EEG stream.

4.5 Signal processing example

In this example, we use prerecorded EEG data to show filter design tools available in Simulink relevant to closed-loop systems. Digital filtering is a necessary preprocessing step in almost every biosignal processing pipeline that requires careful consideration Widmann and Schröger [2012]; Widmann et al. [2015]. FIR filters usually need higher orders than their IIR counterparts to satisfy the same frequency response, but can incur significant time delays. An FIR filter has a linear phase response, which is desirable because, within the band of interest, the filtered signals are delayed by the same amount thereby preventing phase distortions. Time delays can be overlooked when the analysis is performed offline. In real-time systems, however, it is worth considering the use of IIR filters at the cost of introducing small phase distortions. Hopefully, the DSP System Toolbox has tools that allow us to design digital filters precisely quantifying these properties.

We used the pipeline in the left panel of Fig. 4.4 to illustrate time delays introduced by band-pass FIR and IIR filters between 1 Hz and 30 Hz. We used the block *From EEGLAB* to read prerecorded data stored in an EEGLAB-compatible file format Delorme et al. [2011] to simulate an EEG headset. This block reads all the data stored in a file prior to pipeline execution and then emits a multichannel sample on every sample time, where the latter is set automatically as the inverse of the sampling rate declared in the file. We note that this block also guarantees a pipeline execution of at least the sampling period of the EEG by blocking the next execution until that time has elapsed. Although filter blocks can work with multichannel data, in this example we use a *Selector* to process only the Fp1 channel. The motivation for using the Fp1 channel is to use an eye-blink event to illustrate the delays introduced by each filter. Next, we filter data in parallel

using the FIR and IIR filters. For visualization purposes, we also compensate the raw signal by the delay introduced by each filter, concatenate raw and filtered signals and pipe them into a *Scope* block. On the right panel we show the filter design window that pops up if we double click on the *Band-pass* blocks. After designing a filter, we can inspect its frequency response by clicking on the *View Filter Response* button on the top right. We show the frequency response of both FIR and IIR filters in Fig. 4.5.

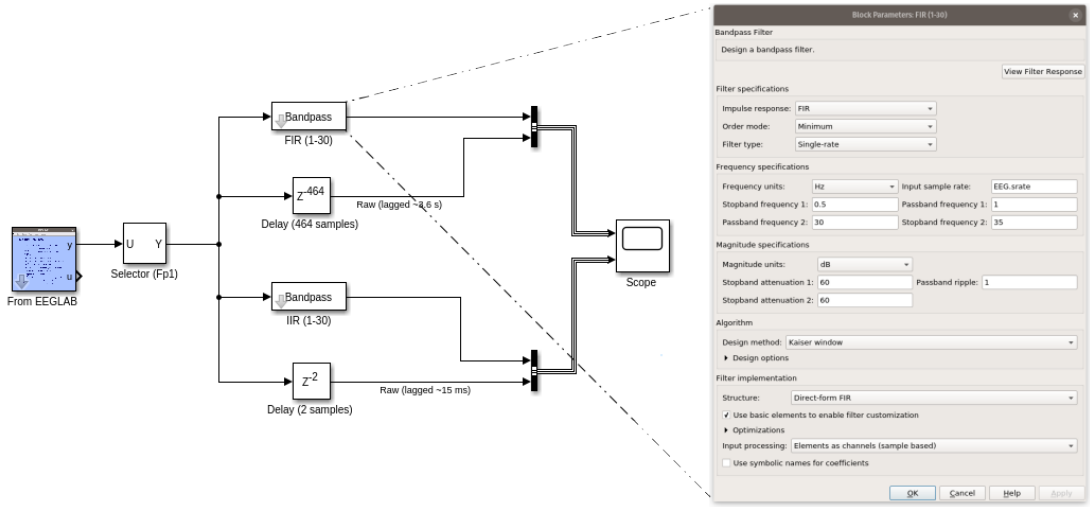


Figure 4.4. Band-pass filtering example in Simulink using blocks from SimBSI and DSP libraries. **Left:** Processing pipeline using prerecorded data. **Right:** Filter design window.

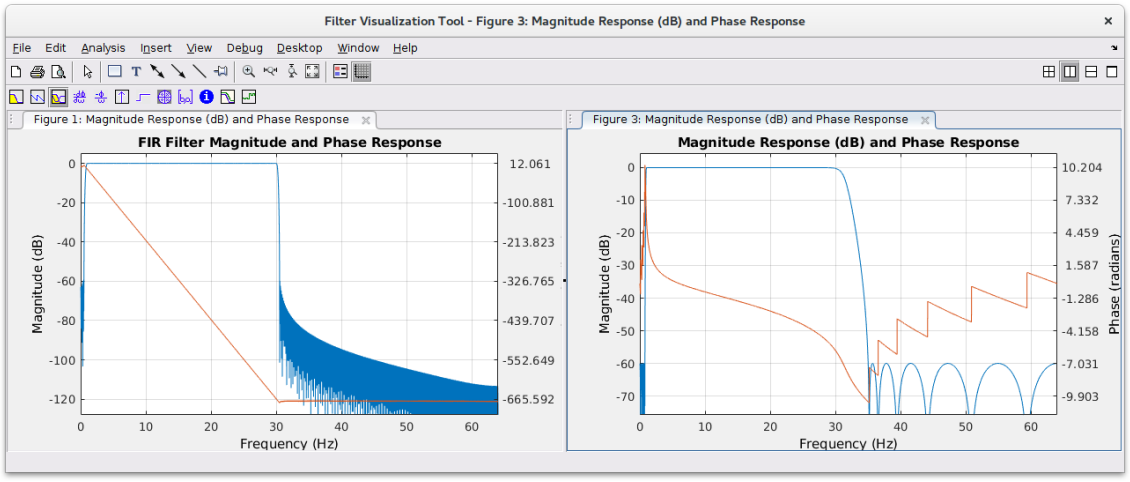


Figure 4.5. Frequency response of FIR and IIR filters used in the example of Fig. 4.4.

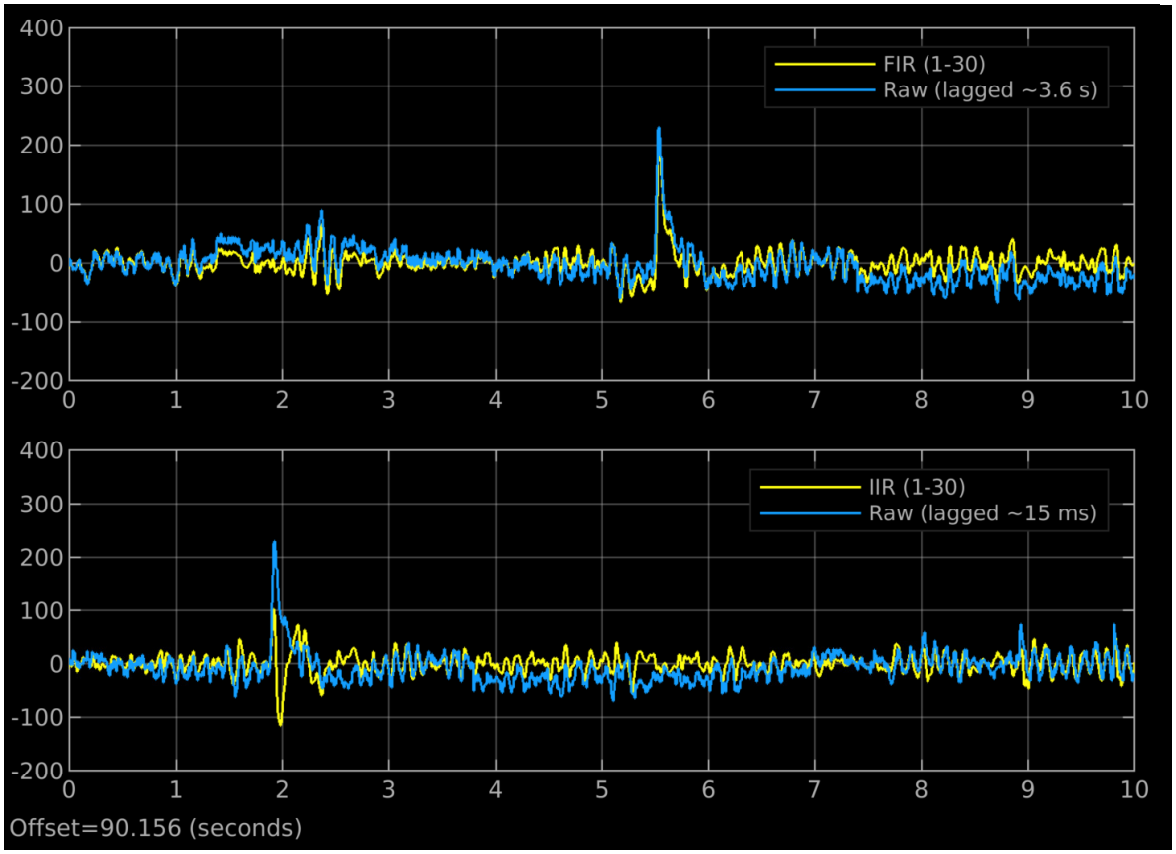


Figure 4.6. 10 sec snapshot of the output of Fig. 4.4 pipeline. The top and bottom panels show the effect of FIR and IIR filters respectively. The eye-blink event was used to illustrate the time delay introduced by each filter.

Fig. 4.6 shows a 10 sec window of the pipeline output in a *Scope* window. Both panels show the filtered and raw signals in yellow and blue respectively. The top panel shows the FIR branch of the pipeline and the bottom panel shows the IIR one. Note that we shifted raw signals in time by the group delay of each filter so that we could visually inspect the distortions introduced. We can see that both filters smooth the data and kept the eye-blink event; again, this is expected because eye movement activity have frequency components higher than 1 Hz. Although the FIR filter did not distort the signal, it introduced a significantly delay of ~ 3.6 sec. Such long delay is prohibitive in a closed-loop system that reacts to brain or behavioral responses in the order of hundreds of milliseconds. The IIR filter exhibited a shorter delay of ~ 31 msec, at the cost of

introducing some phase distortion around the high amplitude eye-blink artifact, which may be a reasonable compromised solution for some applications. We note that, typically, eye-blink artifacts cannot be removed by digital filters alone because they overlap the frequency content of neural signal, thereby calling for more sophisticated artifact removal methods such as the PEB+ algorithm.

4.6 Online EEG source imaging

It is well known that the inverse mapping from EEG scalp sensors to cortical currents is not unique [Michel and Murray, 2012]. Thus, there are many instances where it is more desirable to build a human closed-loop around signals coming from specific brain regions rather than scalp sensors. For example, several studies have pointed out the therapeutic effect of transcranial magnetic stimulation (TMS) to the dorsolateral prefrontal cortex (DLPFC) for treating depression [Pascual-Leone et al., 1996], chronic pain [De Martino et al., 2019], and cocaine addiction [Terraneo et al., 2016], to mention a few. Most TMS protocols, however, are applied in open-loop fashion ignoring ongoing fluctuations in the functional state of the targeted area and this can hinder the efficacy of the treatment [Karabanov et al., 2016]. In this example, we demonstrate an EEG source imaging pipeline for monitoring the activity of several cortical areas in real-time. Then, in Section 4.8 we show a TMS example can be used for building an EEG source-based closed-loop system.

The successful application of EEG-based imaging algorithms depends on the use of data cleaning preprocessing stages that are typically conceived for offline analysis [Artoni et al., 2018]. Recently, we proposed a unifying Bayesian framework, PEB+, that allows for the use of ongoing EEG activity to adaptively segregate cortical source activity into maximally independent components with known anatomical support, while minimally overlapping artifactual activity (see Chapter 2 and [Ojeda et al., 2019]).

In Fig. 4.7, we show an EEG source pipeline where the preprocessing is limited to the common average reference and a 1 Hz high-pass filter. Then, we use the *PEB+* block to minimize the effect of EOG, EMG, and single-channel spikes artifacts while simultaneously estimating EEG source activity in 8003 locations regularly scattered across the cortical surface of the brain. The *PEB+* block relies on the prior deployment of the *Co-register* block to co-register the sensor locations of the incoming EEG data (in this example assumed in the 10-20 system) with a template head model based on the Colin27 head. At initialization time, this block finds the common set of channels between our montage and that of the template, then it selects the corresponding columns of a precomputed lead field matrix[1], which is used down the line by the *PEB+* block for brain mapping. At run-time, the *PEB+* block takes as input a multichannel sample of EEG data and produces the following quantities: $\mathbf{G}$: 8003 cortical source amplitude vector, $\mathbf{V}$: a vector containing the activity of several EOG, EMG and single channel spike artifacts, $\mathbf{Y_c}$: is a sample of the multichannel EEG signal cleaned by projecting out the artifact subspace, λ: an estimate of the common mode sensor noise variance, $\boldsymbol{\gamma}$: a vector estimate of the scale (variance multiplier) of different groups of brain and artifact sources, and logE: the log evidence of the generative model optimized by PEB+, which is a quantity that can be used to measure the quality of the brain mapping procedure at any given time.

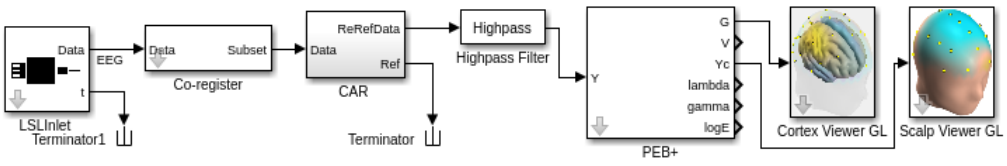


Figure 4.7. Pipeline for online EEG data cleaning, source separation, and imaging using the *PEB+* block and OpenGL viewers.

In our experience, 3D visualizations can severely hinder the real-time performance of a Simulink pipeline. So in this example, to visualize the cleaned EEG and source

[1]The lead field is an overcomplete dictionary of unitary source scalp projections that can be calculated solving Maxwell's equations in a discretized model of the head obtained from MRI data (see [Baillet et al., 2001] for details).

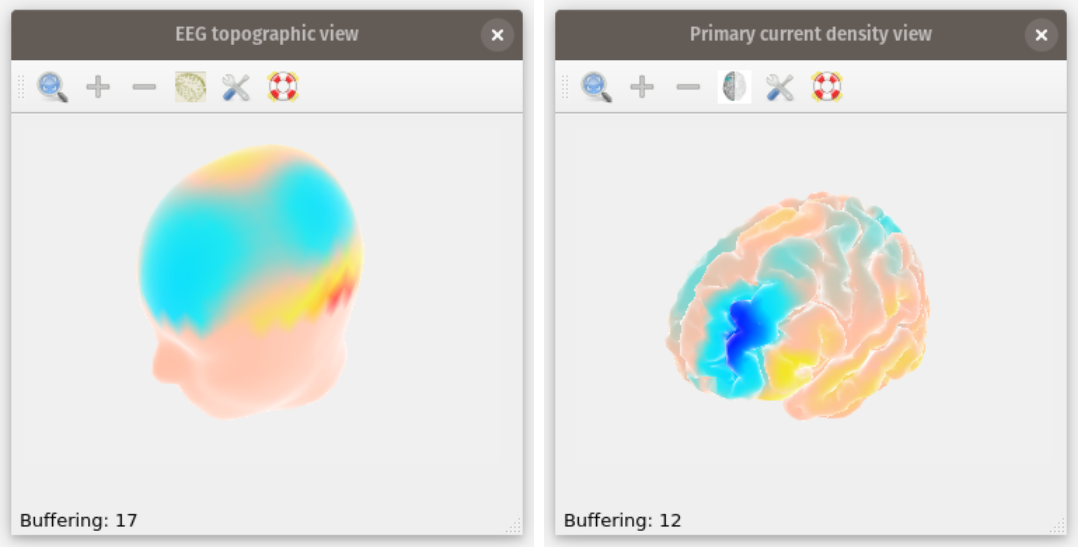


Figure 4.8. Python-based OpenGL real-time scalp and source viewers. **Left:** Scalp viewer [UCSD Neural Engineering and Translation Labs, 2018c]. **Right:** Cortex viewer [UCSD Neural Engineering and Translation Labs, 2018b]. Follow the link to see a movie of a live visualization using these apps https://www.dropbox.com/s/rkkz5ywj7m8xjyb/rt_esi.mp4?dl=0.

estimates produced by the *PEB+* block we use two freely available OpenGL-based apps written in Python, the ScalpViewer [UCSD Neural Engineering and Translation Labs, 2018c] and the CortexViewer [UCSD Neural Engineering and Translation Labs, 2018b]. The blocks *Cortex Viewer GL* and *Scalp Viewer GL* forward the data they receive to LSL which are then rendered externally by the Python apps. In the specific case of the *Scalp Viewer GL* block, it also linearly extrapolates the voltages from the sensor locations to the rest of the scalp so that we can render EEG topographies as a continuous field. We note that, in addition to LSL, the Python apps can consume data using the TCP/IP protocol, so in application where the data need to be inspected remotely, we can simply replace the *GL* blocks in the pipeline by standard *TCP/IP Send* blocks from the Instrument Control Toolbox.

4.7 Stateflow cognitive task design

In addition to having a fully vetted graphical environment for signal processing, of which we have full control, ideally one would like to have an experimental environment that exhibits these same features. In this section, we explore the use of Stateflow for the design of cognitive tasks within the Simulink environment. Performing the experimental design within Simulink is advantageous for cross-species studies because we can implement the core logic of the task once, and then use it with species-specific data acquisition and stimulus presentation devices. Stateflow is a visual environment where we can design sequential decision logic triggered by subject's behavioral responses based on state machines and flow charts [MathWorks, 2018b]. Furthermore, Stateflow allows us to visualize state transitions while the subject is playing the task, which is especially useful for debugging in the earlier stages of the experimental design.

In this section we use a special type of Simulink block called Stateflow *Chart*. A *Chart* is composed of a collection of states whose transitions can be triggered by internal or external signals. Inside a *Chart*, blocks represent states while state transitions are represented by arrows. Each state has a name (mandatory) followed by any number of instructions written in MATLAB language, this is what the state does and is called "state action". State actions can be of three types depending on when we want the code to be executed: 1) *entry* – right when the state becomes active, 2) *during* – during the time that the state is active, and 3) *exit* – right before transitioning out of the state. In this example we only use entry actions because there is no obvious reason to do otherwise but we encourage practitioners to check Stateflow's documentation for other options and use cases. A state can have any number of transitions which are then triggered in order of priority by combining logical and temporal statements. For example, in Fig. 4.9 **B**, the fourth transition of the state *WaitForResponse* is triggered after 1.8 sec of being in that state and a Go image was displayed.

Fig. 4.9 shows an example of a Go/NoGo task designed in Stateflow. Panel **A** show the Simulink system. The blocks enclosed in the purple areas represent task parameters (Go stimulus probability and number of trials) and the acquisition of subject's behavioral responses. In this example, human subject's key presses are captured by a key logger app written in Python that runs in the background. The logger app forwards the key codes to LSL, which are then read in by Simulink using the *Key Logger* block (which internally uses the *LSL Inlet* block). The blue area groups species-independent blocks for the probabilistic presentation of Go/NoGo images and the trial logic. The orange area groups visual and auditory stimuli presentation blocks as well as task and performance measures (trial counter, correct and incorrect trials, subject's key presses, etc) that may be useful to the experimenter.

Fig. 4.9 **B** shows the *Trial Logic* block implemented in a *Chart*. From top to bottom, the entry point to the system is the *TaskInit* state, this is where we initialize variables and prepare for subsequent actions. To transition out of the entry point to the *NewTrial* state, and thereby start the task, we wait for the user to press the Enter key (code 65293). In *NewTrial* we increment the trial counter and set *show*=1, when this signal rises it triggers the *Go/NoGo Image Presentation* block outside the chart to randomly select the image that will be immediately shown on the screen. After 200 msec we transition to the *WaitForResponse* state and set *show*=0, which has the effect of removing the image and displaying a blank screen. *WaitForResponse* has four possible transitions: 1) to *CorrectTrial* if *response* is equal to 32 (space bar key code) and was a Go image (correct go condition), 2) to *ErrorTrial* if the response is 32 but was a NoGo image (incorrect go condition), 3) to *CorrectTrial* if there was no response for 1.8 sec and was a NoGo image (correct withholding condition) and 4) to *ErrorTrial* if there was no response for 1.8 sec and it was a Go image (incorrect withholding condition). In the *CorrectTrial* and *ErrorTrial* states we activate the respective correct or erroneous auditory feedback tone for 100 msec and then we transition to *EndOfTrial*. If we reached the maximum

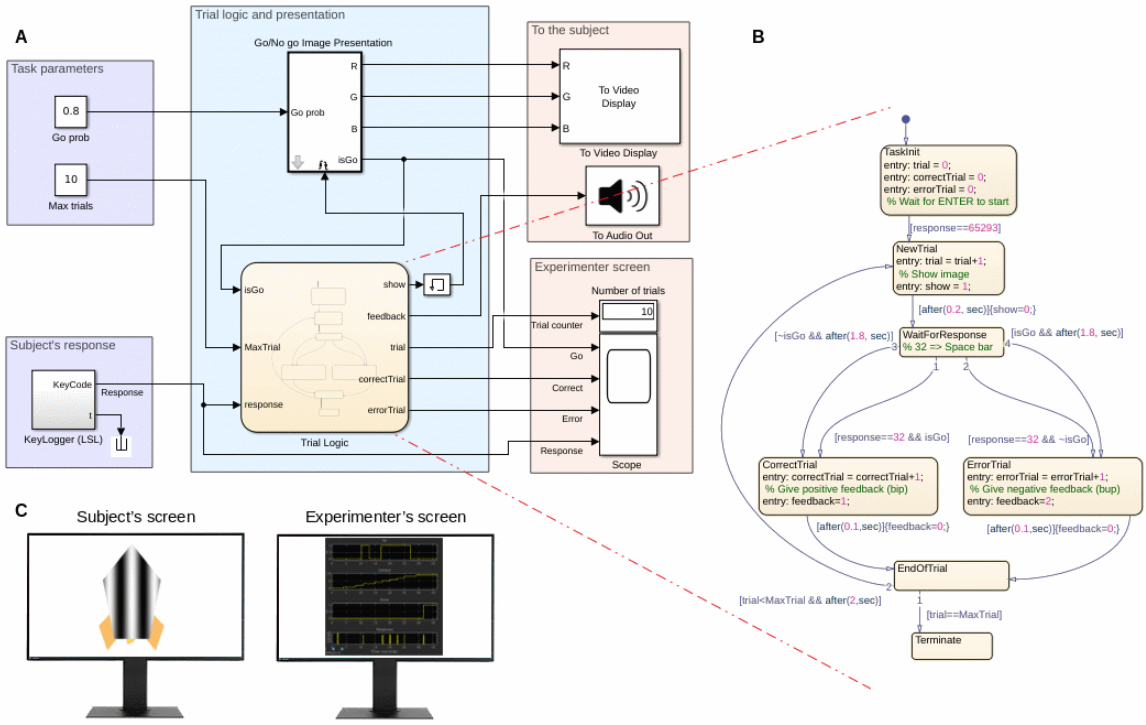


Figure 4.9. Stateflow experimental design.

number of trials we terminate otherwise we wait for 2 sec and start a new trial.

Panel **C** represents a dual monitor setup in which the left screen can be used for subject's stimulation while the right one could face the experimenter in a control room. Both screen can be controlled by Simulink and all the relevant signals flowing through the system can be saved in a Matlab file using the *To File* block or sent to LSL through *LSL Outlet* blocks and then saved in the *.xdf* file format using the *LabRecorder* app [UCSD Swartz Center for Computational Neuroscience, 2018a].

4.8 EEG-based closed-loop neurostimulation

TMS has become a popular technique for studying the relationships between brain and behavior [Thut and Pascual-Leone, 2010] through noninvasive interventions both in healthy human subjects and patients. As such, TMS have been proposed as a valuable experimental and therapeutic tool for closed-loop neuroscience [Zrenner et al., 2016].

For instance, it has been shown that stimulation guided by the phase and frequency of intrinsic EEG rhythms can result in different types of long-term plasticity [Zrenner et al., 2018]. Using this as a motivation, in Fig. 4.10 we demonstrate the relative simplicity of implementing a TMS closed-loop human BCI pipeline in Simulink.

At a higher abstraction level, the pipeline is composed of four major submodules for: 1) EEG data acquisition, 2) occipital alpha phase estimation (brain state estimation), 3) control and stimulation, and 4) visualization. The pipeline is designed to acquire EEG samples from LSL and continuously monitor the power in the alpha band of the Oz channel. When the power in the alpha band is greater than other frequency components within 1 Hz and 40 Hz for at least 1 second, a single TMS pulse is triggered phase-locked to the positive peak of the EEG signal in Oz. Since we build on standard DSP blocks, the pipeline allows to flexibly change the monitored channel, frequency, phase of interest, as well as the stimulation pattern sent to the stimulation hardware. Figures 4.10-4.14 show in greater detail the different components of the system.

In this example, we send trigger pulses from Simulink to a MagVenture MagPro TMS machine using a PulsePal device [Sanders and Kepecs, 2014]. PulsePal is an inexpensive and low-latency pulse generator that has been used in closed-loop electrophysiological experiments to deliver auditory and optogenetic stimulation to mice during a Go/NoGo task [Pi et al., 2013]. Furthermore, PulsePal has been recently integrated into Open Ephys. We used PulsePal's open-source C++ DSK to write an s-function and interface it with Simulink, which we encapsulated into the *To PulsePal* block. We note that by having a common software stack, it becomes relatively easy to translate between human and animal closed-loop experiments, keeping most parts intact and swapping out species-specific components only, e.g., EEG by LFP signals and TMS by optogenetic stimulation hardware, and so on.

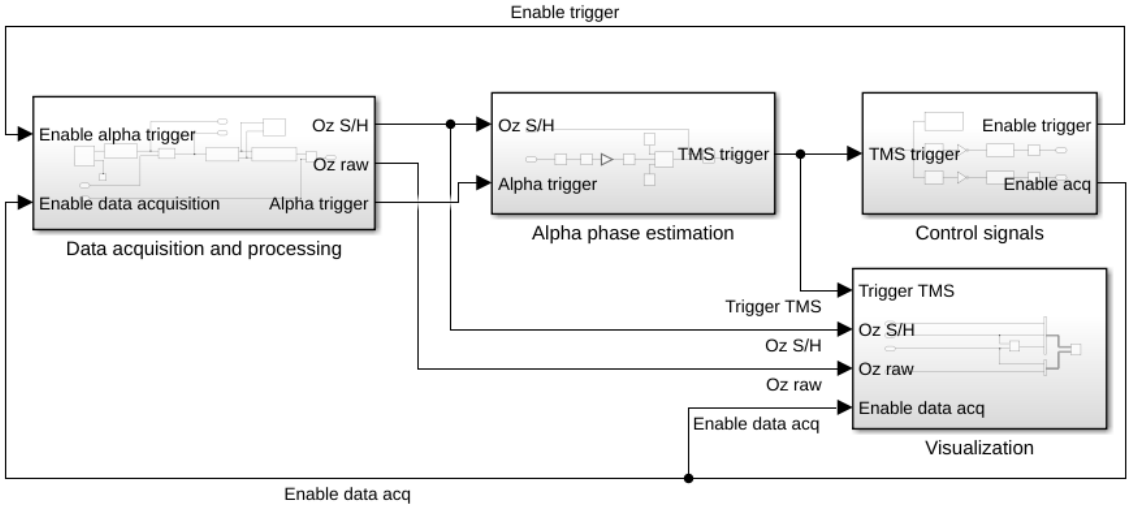


Figure 4.10. Closed-loop BCI with TMS stimulation composed of four major modules: 1) Data acquisition, 2) EEG alpha wave phase estimation, 3) Control and stimulation, and 4) Visualization. The system is designed to acquire EEG data through LSL and trigger a TMS pulse if an alpha wave is detected at the Oz channel and it is stable for at least one second. The stimulation pulse is locked to the positive peak of the EEG signal, which a that point should consist of mostly alpha rhythm activity. Each submodule is explained in the panels below.

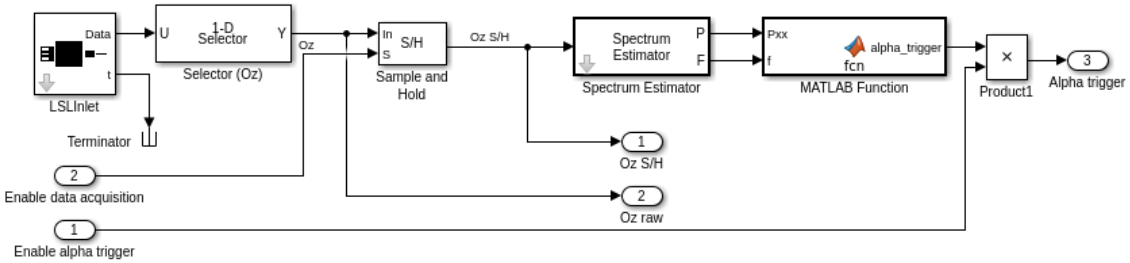


Figure 4.11. Data acquisition submodule: 1) The *LSLInlet* pulls EEG samples from LSL at the sampling rate of the EEG acquisition device. 2) The *Selector* block selects the subset of channels we are going to work with, which in this case is Oz. 2) The *Sample and Hold* block enables EEG samples to pass through while the TMS pulse is not active, otherwise, it holds the last sample acquired before a TMS pulse and resumes sampling once the artifacts are no longer present in the measurement channel. 4) The *Spectrum Estimator* block computes the Welch (or filter bank) power spectral density estimates on 1 sec windows with a 50% overlap. 5) User-defined function that sums the power in $\delta, \theta, \alpha, \beta, \gamma$ bands and outputs 1 if the maximum power is in α, otherwise, the output is 0.

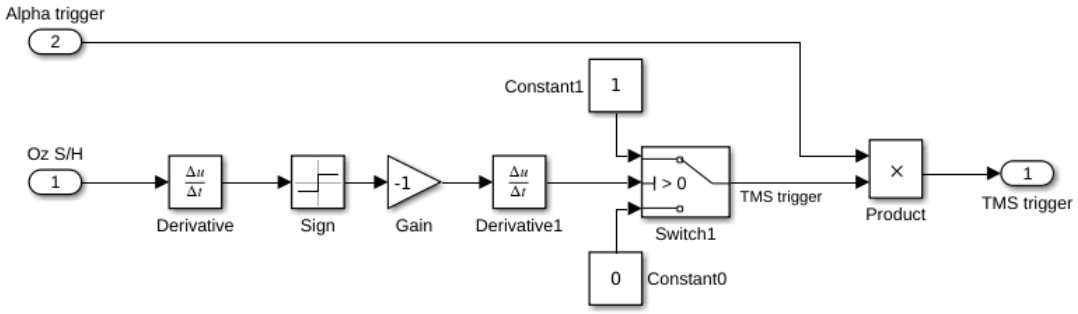


Figure 4.12. Submodule for triggering a TMS pulse locked to the positive phase of the alpha wave. The bottom branch produces a Kronecker delta at the positive peak of the Oz channel, say $y(t_k)$, implementing the expression $\mathrm{diff}(-\mathrm{sign}(y(t_k) - y(t_{k-1})))$. Note that to phase-lock to the trough of the wave we just flip the sign of the unitary gain above.

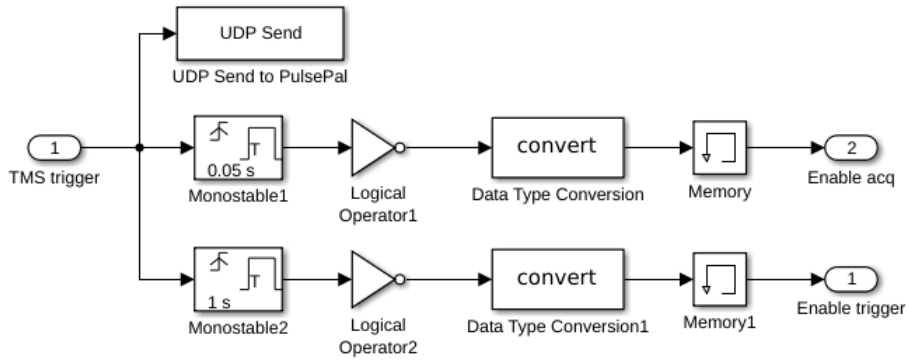


Figure 4.13. Control submodule. Once we have determined that a TMS pulse should be delivered, we send out the stimulation trigger by the serial port and use two *Monostable* blocks to 1) disable the data acquisition for a time span approximately equal to the duration of the TMS artifact and 2) disable stimulation for a period of time. This can be reconfigured depending on the stimulation protocol, e.g., single pulse, burst, and so on.

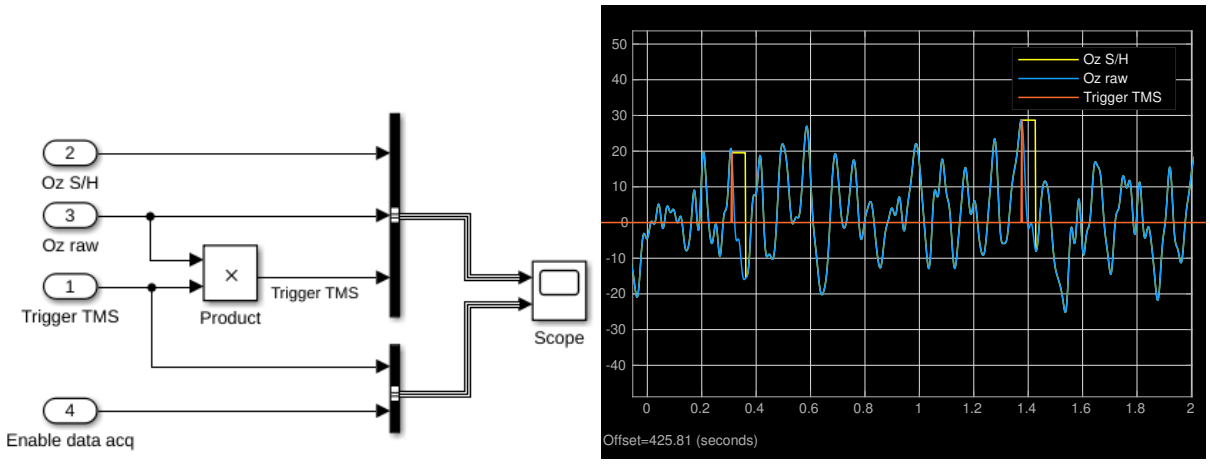


Figure 4.14. Left: Visualization submodule. **Right:** Excerpt of 2 sec data flowing through the system.

4.9 Animal experimental environment

Our last example blends most elements explained so far (with the exception of human EEG imaging) in the creation of an experimental environment for rodents. The environment consists of two major components: 1) a behavioral box and 2) an electrophysiology data acquisition system. Although this environment will be described in detail in a separate publication, next we proceed to briefly explain how we used elements in the SimBSI library and auxiliary apps to integrate Stateflow task design, behavioral and electrophysiology data acquisition, stimulation and post hoc data analysis.

4.9.1 Behavioral box

The behavioral box consists of a custom built chamber where animals are placed to perform a cognitive task while we record their behavioral responses and electropysiological signals (see Fig. 4.15 **A**). The chamber has a display where visual stimuli can be presented. In front of the display there are 5 nose ports, each of which is equipped with infrared (IR) sensors for collecting animal choices and a liquid reward delivery mechanism. There are speakers placed on each side of the display for delivering auditory feedback and house lights glued to the top face of the chamber, which are often used to indicate to the animal

the start of the task or to reinforce negative feedback in the form of short bursts of 10 Hz flashing light.

The display, IR sensors, reward mechanism, auditory feedback, and house light are controlled by a Simulink process running on a Raspberry Pi 3 model B+ (Pi for short) that is attached to the outer face of one of the walls of the box (see Fig. 4.15 **B**). The Pi interfaces with sensors and actuators via the GPIO pins connected to a custom-built power electronics control board. Before deployment to the behavioral box, we assign an IP address to the Pi and configure it to connect to a dedicated LAN (typically wireless), this way we can have several boxes in the lab and treat them as self-contained mobile experimental environments.

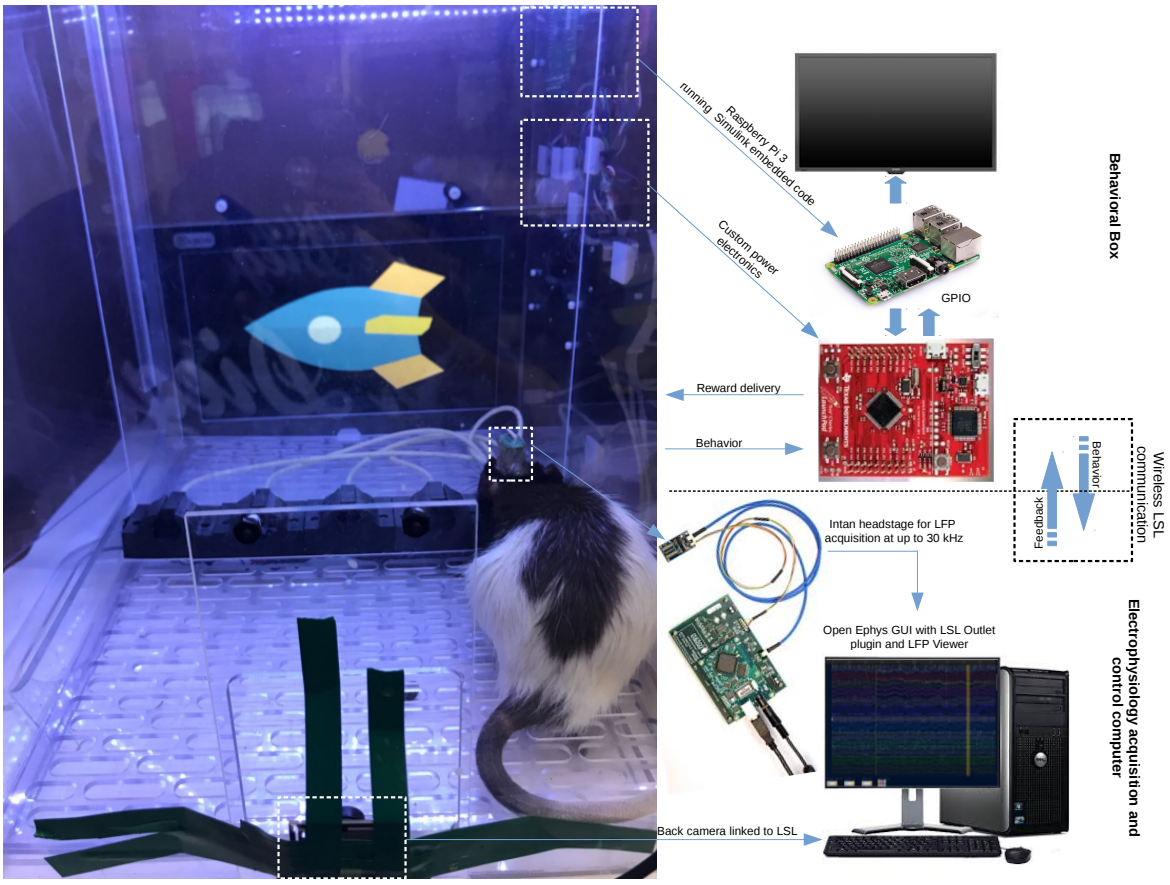


Figure 4.15. Animal experimental environment.

Once the Pi is fully configured and attached to the box, it can be programmed

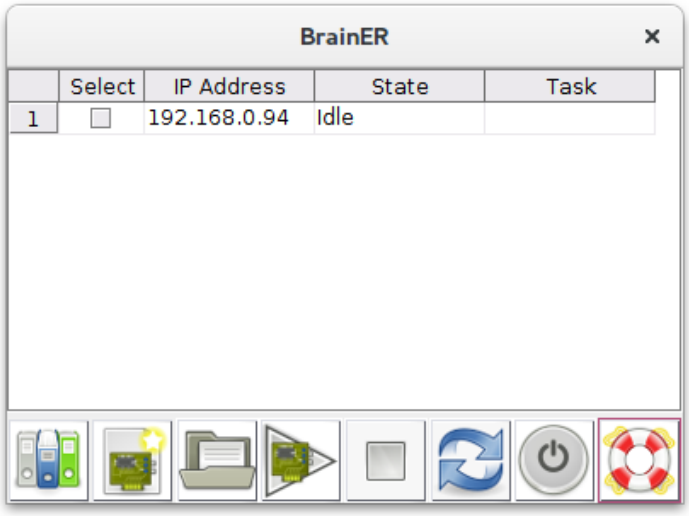


Figure 4.16. The BrainER GUI allows us to manage the software that runs on Pis that control several behavioral boxes that we may have in the lab. Once the GUI is first launched, it searches the network for Pi devices connected between IP addresses 192.186.0.90-150.

using a Simulink session running on a remote computer connected to the same network. Common cognitive tasks such as the one shown in Section 4.7 can be coded using Stateflow and Simulink blocks in the host computer, simulated, and then deployed to the targeted Pi using MATLAB Coder. MATLAB Coder compiles a Simulink program into C, which is then optimized and compiled specifically for Pi's hardware. The resulting binary can be launched on the Pi remotely from the host computer and run as a real-time standalone application. Once the task is compiled, the binary stays permanently in the Pi's SIM card. To access compiled tasks stored on the Pi remotely without using Simulink, we have created a MATLAB GUI called BrainER[2] (see Fig. 4.16). With BrainER we can select one or several Pis and reconfigure, stop, and start tasks on them.

4.9.2 Electrophysiology data acquisition system

The animals that participate in the tasks administered through the behavioral box are usually implanted with intracranial electrodes. To acquire LFP data we use Intan

[2] https://bitbucket.org/neatlabs/brainer/wiki/Home

headstages connected to a RHD2000 acquisition board which is controlled using the Open Ephys software. To synchronize the LFP data with the control and behavioral signals coming in and out of the behavioral box we use LSL. As we mentioned in Section 4.4, we use a *LSL Outlet* plug-in for Open Ephys to forward to LSL each LFP sample acquired. To enable LSL communication on the behavioral box, we used the read/write GPIO blocks in the Raspberry Pi support library to create new blocks that also forward to LSL the signals they read in or write out respectively.

Since LFP, behavioral, and control signals can be accessed via LSL, we can save the data generated on each session using the LabRecorder app, and analyze them offline in MATLAB using MoBILAB [Ojeda et al., 2014] and EEGLAB [Delorme et al., 2011] toolboxes and custom scripts. Furthermore, in experiments where we need to close the loop, we can implement an additional Simulink pipeline that reads from LSL the LFP data to monitor a target brain state and trigger a stimulation device similar to what we did in Section 4.8.

4.10 Conclusions

In this chapter, we have developed SimBSI, an open-source Simulink library for the rapid prototyping of brain signal interfaces. We designed the library to achieve the following design principles: 1) easy to adopt by users of different programming backgrounds, 2) transparency of data processing, 3) multiplatform, and 4) flexible data acquisition. Most of these design principles are achieved by using a mature signal processing environment such as Simulink, powered by an intuitive graphical programming language. Simulink programs are multiplatform and can be deployed as standalone applications that run on standard or embedded hardware such as a Raspberry Pi and Arduino. We extended the data acquisition capabilities of Simulink by implementing LSL streaming blocks. We used several examples to demonstrate the capabilities of the library for implementing cross-

species BCIs, ranging from simple signal processing, EEG source imaging, task design, to closed-loop neuromodulation. Furthermore, we demonstrated that a sophisticated experimental environment for animals is feasible and relatively straightforward to develop within Simulink using SimBSI.

4.11 Acknowledgments

This research was supported by NIMH Training Fellowship in Cognitive Neuroscience (AO) grant MH020002, UC San Diego Chancellor's Research Excellence Scholarship (JM, AO), and UC San Diego School of Medicine start-up funds (JM). The PEB+ algorithm is copyrighted for commercial use (UC San Diego Copyright #SD2019-810) and free for research and educational purposes.

This chapter is currently being prepared for submission for publication of the material. I am the primary investigator and author of this material along with co-authors Nathalie Buscher, Pragathi Priyadharsini Balasubramani, Vojislav Maric, Dhakshin Ramanathan, and Jyoti Mishra.

Chapter 5

Towards source-based identification of attention switch in an unconstrained computerized task

5.1 Introduction

Cognitive control is a fundamental human ability that allows us to flexibly pay attention to and act upon goal-relevant information, while further suppressing irrelevant distractions. Higher cognitive skills such as memory, learning, and task planning crucially dependent on it [Fortenbaugh et al., 2017; Gazzaley and Nobre, 2012]. Since sustained attention is a plastic brain function, the overuse of distracting technologies can deteriorate its deployment at times when we need to focus on a task, thereby leading to reduced productivity [Ophir et al., 2009; Ziegler et al., 2015]. It has been estimated that U.S. children between 8 and 18 years old spend an average of 9 hours of their day consuming media in the form of television, Internet, email, video games, social networks, and interaction with mobile devices [CSM, 2015]. Moreover, despite the wide availability of pharmacologic therapies, Akinbami et al. [2011] found in a similar sample that the number of individuals diagnosed each year with attention deficit disorders (ADD) is steadily growing.

Here, we develop noninvasive mobile brain-computer interface (BCI) neurotechnology to address this problem. A BCI can be used to encourage the subject to skip

distractions while reinforcing the use of brain circuits that support sustained attention. Although attention has been studied within the constraints of the lab [Fortenbaugh et al., 2017], its biological underpinnings as a whole, acting during unconstrained daily-life activities remain poorly understood. Towards realizing our vision of a neurotherapeutic BCI, in this chapter, we use the brain imaging methods developed in earlier chapters to study the internally driven deployment of attentional resources in healthy adult subjects performing a self-paced schoolwork-like computerized task. In particular, we focus on characterizing the brain network dynamics during transitions from periods of sustained attention to distraction.

Several studies have shown that periods of lower sustained attention are characterized by high response time (RT) variability and frequent errors on task [Esterman et al., 2013; Fortenbaugh et al., 2015]. These metrics can be obtained on trial-based continuous performance tasks (CPTs) where individuals discriminate task-relevant target information from task-irrelevant non-targets [Conners et al., 2003]. A recent study acquired data from thousands of individuals across the lifespan on a CPT paradigm and showed that when subjects were attentive (or "on-task"), they exhibited lower RT variability and higher discrimination ability as opposed to when they were distracted (or "off-task") [Fortenbaugh et al., 2015]. Moreover, RT variability has been proposed as a bio-marker for ADD [Di Martino et al., 2008]. Building upon these results, here we propose to define periods of putative sustained attention and distraction from the variability of keystroke and mouse click events produced by the subject during a self-paced schoolwork computerized activity (see task description in Section 5.2.1).

To characterize the fluctuations in brain dynamics that are predictive of incoming distracted behaviors at a millisecond time scale, we collect EEG data in sync with behavioral events. In addition to its excellent temporal resolution, EEG is appealing for BCI interventions because it is noninvasive and available as low-cost mobile technology [Bateson et al., 2017; Kumari et al., 2017]. Traditional EEG-based BCIs are implemented

using features obtained in sensor space [Lotte et al., 2018]. However, since the inverse mapping from EEG sensor voltages to cortical currents is not unique [Lopes da Silva, 2013], sensor-space features constitute distorted descriptors of the underlying brain processes of interest and are often confounded by other non-neural (ocular/muscular) responses. Thus, despite impressive advances in signal processing and machine learning [Congedo et al., 2017; Lotte et al., 2018], current BCIs interventions offer marginal therapeutic value [Lee et al., 2013]. To overcome these limitations, in this chapter, we use the EEG source imaging algorithm PEB+ developed in Chapter 2.

PEB+ extends the PEB framework previously proposed for EEG source imaging [Henson et al., 2011; Wipf and Nagarajan, 2009] in two ways. First, it augments the standard generative model of the EEG with a dictionary of empirically-obtained stereotypical artifact scalp projections. Second, it uses an anatomic parcellation of the cortical surface to constraint the source estimates to be sparse in the number of areas that are active at any given time. These two elements have the property of inducing the segregation of the source activity into a few maximally independent components (ICs) with known anatomical support, while EOG, EMG, and single-channel spike artifacts are also segregated to their respective ICs. Furthermore, while other imaging algorithms are limited by their computational footprint, PEB+ is capable of real-time performance and is resilient to fluctuations in the variance of the sensor noise corrupting EEG samples (see Chapter 3).

Although central to our approach, source estimation alone cannot unveil how different brain regions dynamically interact to support attentive and distracted behaviors. To that end we use source time series to estimate source connectivity. Popular approaches for estimating EEG source connectivity use parametric models based on strong biophysical [David et al., 2006; Kiebel et al., 2009] or statistical [Giraldo et al., 2010; Yamashita et al., 2004] assumptions. These approaches tend to give reasonable results when used to analyze event-related brain responses. In this chapter, however, we characterize source connections

on an ongoing basis, therefore, without prior knowledge of what cortical regions may be active or what type of temporal dynamics are expected to arise (i.e., steady-state, non-stationary, linear or nonlinear dynamics), the assumptions of the parametric models are hard to justify. To circumvent these issues, Lizier et al. [2008] proposed to compute the transfer entropy (TE) between the electrical activity of pairs of brain areas, conditioned on the activity of all the other areas, as a non-parametric measure of their effective (causal) connectivity (EC). Furthermore, it has been shown that TE-based EC is sensitive to ongoing linear and nonlinear brain dynamics [Wibral et al., 2014].

A data-driven characterization of the EC of large-scale cortical networks that correlate with attention switches in an unconstrained task such as ours calls for a comprehensive statistical analysis that is beyond the scope of this chapter. As a proof of principle, here we focus on testing the presence of salience network (SN) activity during the putative transitions from on-task to off-task periods. The SN connects the anterior insula (AI) and the anterior cingulate cortex (ACC). Based on results from fMRI data, Menon and Uddin [2010] proposed the AI as a hub that mediates the dynamic interactions between different large-scale networks and is implicated in attention switches driven by external and internal events. In their model, Menon and Uddin [2010] postulate that the AI is sensitive to salient events, and that its key function is to identify such events for additional processing, while the coupling with ACC facilitates rapid access to the motor system, thereby triggering behavioral changes when needed. They also showed that the activity of the SN is more lateralized to the right (rAI). To the best of our knowledge, the causal interplay between nodes in the SN during unconstrained tasks involving internally-driven attention switches has not been yet demonstrated on EEG data. So, to this end, here we use these previous results to hypothesize that within 2 seconds around the transition from on-task to off-task, there will be a causal influence from rAI to ACC. If we can prove this hypothesis, we will move one step closer to understanding the mechanisms of attention regulation in the brain at a millisecond resolution.

The rest of the chapter is organized as follows. In Sections 5.2.1-?? we describe the task, the behavior and EEG data processing, feature extraction, and our model learning approach. In Section 5.3 we present our results of learning subject-level and group-level classification models that map source-resolve features to on-task and off-task states. We finalize with a discussion in Sections 5.4 and our conclusion in ??.

5.2 Materials and methods

5.2.1 Task description and experimental setup

A total of 25 subjects participated in the study, 46.43% male and 53.57% female, ages 23.3 +/- 5.7 years old. All participants were informed of the nature of the experiment before signing a consent form. The experiment was approved by the IRB committee of the University of California San Diego.

The task consisted of sitting in front of the computer for one hour to solve high school level problems combining self-chosen modules of math, critical reading and writing. The subjects were instructed to work through the problems at their own pace. We did not record their scores. As in any real-world computerized task, we allowed them to self-choose when to be on-task and when to be distracted and move away from the task (see Fig. 5.1).

We measured subject's behavior by capturing the keystroke and mouse click events that they generated throughout the task. To this end, we implemented a secure key and mouse logger app that ran in the background without user intervention. To co-register the behavioral events with the EEG signal we used the Lab Streaming Layer (LSL[1]) library. Each keyboard or mouse event captured received a time stamp by sending to LSL the characters '1' or '2' respectively, thereby obfuscating the exact identity of those events. We acquired the subject's EEG at 500 Hz using an LSL-compatible mobile amplifier (Smarting, mBrainTrain, Belgrade, Serbia) and a 24-channel montage placed according to the 10/20

[1]https://github.com/sccn/labstreaminglayer

system. We recorded event markers and EEG data into an EEGLAB-compatible *.xdf* file using the *LabRecorder*[2] app.

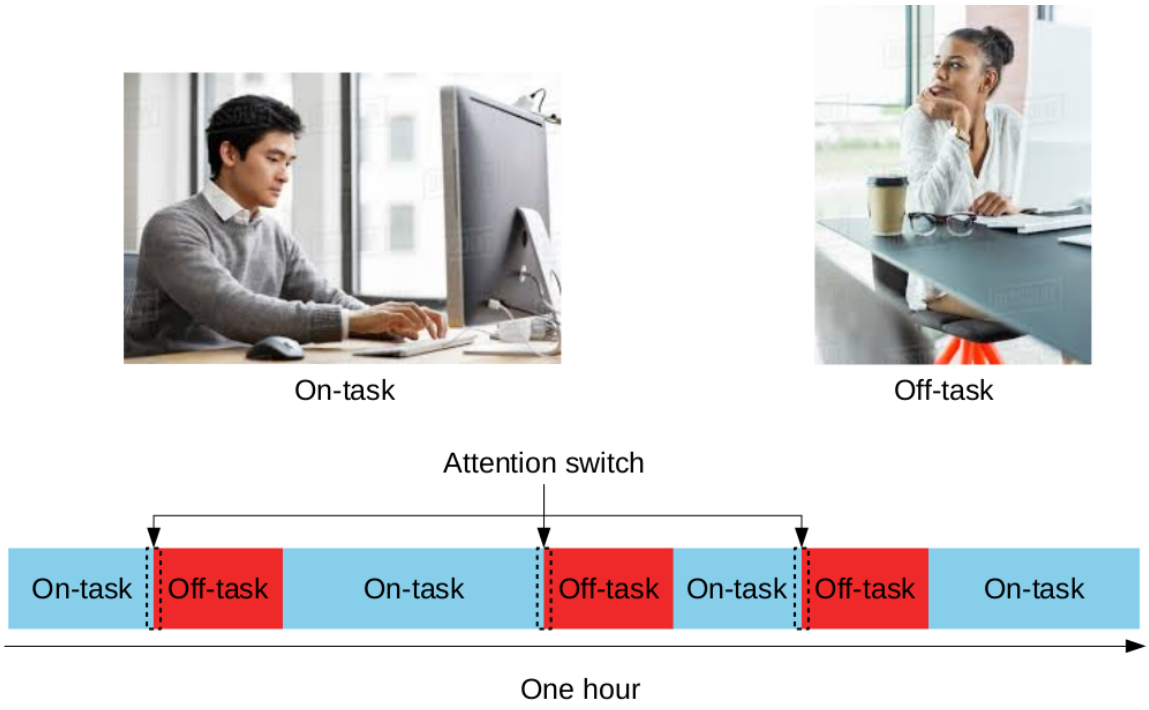


Figure 5.1. Self-paced schoolwork-like task. The subjects were instructed to solve problems combining self-chosen modules of math, critical reading, and writing for one hour. They were allowed to take as many breaks they wanted with the only restriction of moving away from the computer during breaks. The blue and red blocks represent putative on-task and off-task periods respectively, which we identified using the variability of the subject's interaction with the computer. We used EEG data one second after and before putative transitions from on-task to off-task periods to study the network correlates of attention switch.

5.2.2 Identification of on-task/off-task epochs

We defined periods of putative sustained attention (on-task) and distraction (off-task) using the variability of keystrokes and mouse click events produced by the subject during the task. We used a subject-specific 95 percentile of RT to indicate an off-task switch (see left panel of Fig. 5.2 and inset histogram). We defined attention switch events as the last behavioral event before an off-task period. In Fig. 5.2 we show behavior and

[2]https://github.com/labstreaminglayer/App-LabRecorder

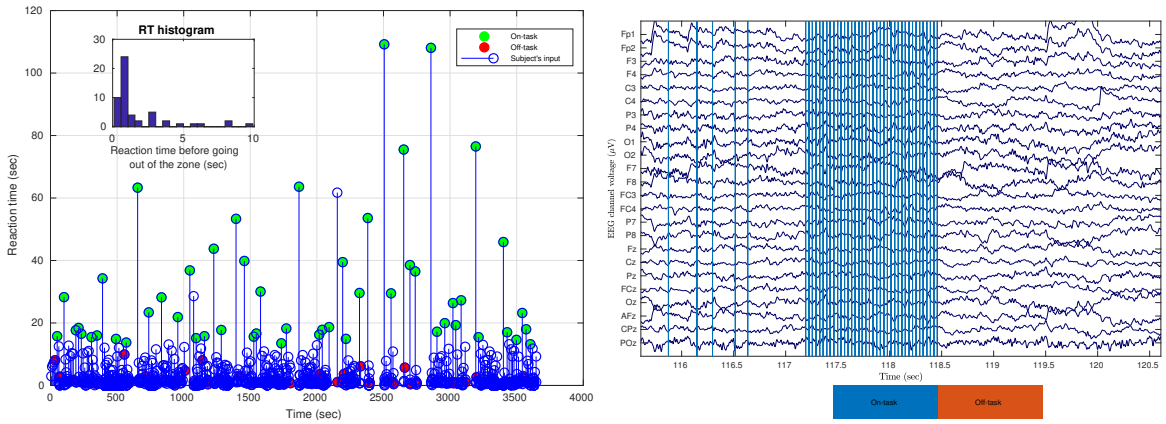


Figure 5.2. Identification of attention switch events as the transition from on-task to off-task epochs. **Left:** Behavior of one participant (keystroke and mouse click events) captured by the logger app. The inset histogram shows the distribution of RTs, where the 95 perentile was used to identify the off-tasks periods. **Right:** Co-registered behavioral events and EEG signal for a portion of the session exhibiting an attention switch from on-task to off-task states.

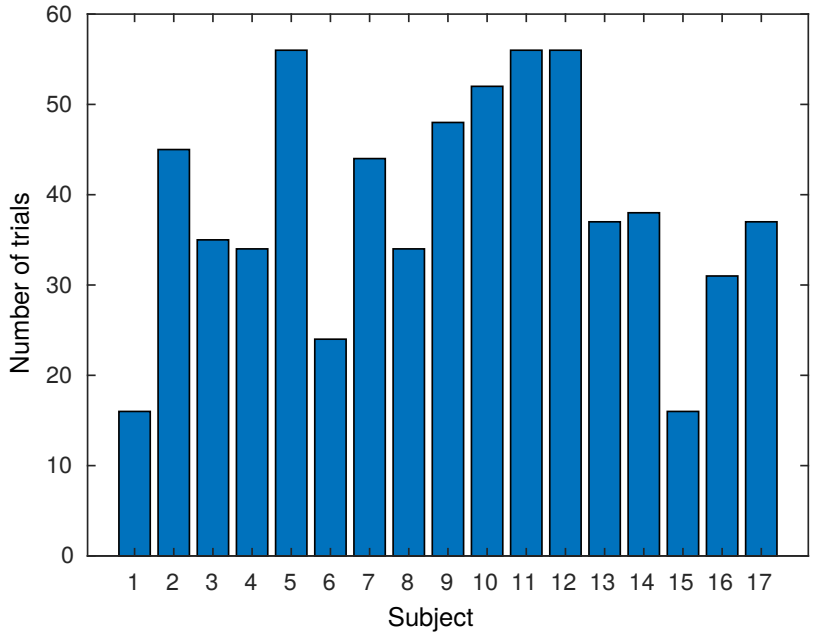


Figure 5.3. Number of trials collected by subject.

EEG data from one participant. Fig. 5.3 shows the total number of trials collected by subject.

5.2.3 EEG preprocessing

We processed the EEG data in MATLAB (R2018b The MathWorks, Inc., USA) using the EEGLAB toolbox [Delorme et al., 2011]. The preprocessing consisted of a 0.5 Hz to 50 Hz band-pass zero-lag FIR filter and re-referencing to the common average channel. We note that this choice of filter is convenient for offline analysis as it doesn't shift EEG samples with respect to event markers. In an actual BCI, however, we would use a minimum lag IIR filter and shift behavioral events according to the group delay of the filter. We visually inspected the data and removed 8 subjects from the analysis because they exhibited more than three faulty channel connections throughout the task. Next, we collected trials consisting of one second before and after the attention switch event and manually removed those excessively contaminated by artifacts. Trials contaminated with EOG and EMG artifacts were kept because those artifacts were dealt with in the source estimation stage by the PEB+ algorithm. Since we didn't have access to a digitizer, we used channel labels to look up the columns of a precomputed lead field matrix defined on the four-layer (scalp, outer skull, inner skull, and cortex) "Collin27" template head. The lead field matrix was precomputed for a 339-channel superset of the 10/20 system using the OpenMEEG toolbox [Gramfort et al., 2010].

5.2.4 Source power estimation

We used the PEB+ algorithm to estimate single trial EEG source time series resulting in arrays of 8003 cortical sources by 2 sec by trial by subject. The PEB+ algorithm uses sparsity constraints to automatically segregate the cortical activity into a sparse number of independent clusters of sources active at any given time. The group-sparsity constraints were enforced using 68 ROIs defined in the Desikan-Killiany atlas [Desikan et al., 2006] as in Chapter 3. After source estimation, we calculated the source ROI power time series by summing up the square of the activity within each ROI.

5.2.5 Source effective connectivity estimation

We used the ROI power time series to calculate the time series of the functional connectivity between every pair of ROIs. Functional connectivity was characterized by TE, which we computed on a sliding window of 100 msec with a 50% overlap. We defined the strength of each causal connection as the maximum TE value across several lags. Furthermore, to factor out common influences from other ROIs, we conditioned the TE on the activity of those. We used Kraskov's TE estimator [Kraskov et al., 2004] as implemented in the Java Information Dynamics Toolkit (JIDT) [Lizier, 2014].

5.3 Results

Although we estimated the EC between all pairs of connections, which constitutes a wealth of information in on itself, in this section, we followed a hypothesis-driven analysis limited to the activity of the SN. In particular, we tested whether there is a causal influence of the rAI over the ACC (rAI→ACC) within the 2 sec window relative to the putative attention switch event. If there is, this would suggest that the detection of salient events is related to the observed transition in subject's behavior from on-task to off-task periods. To this end, first, we investigated whether there is network activity time-locked to the putative attention switch event. In Fig. 5.4, we show the time series of rAI→ACC connectivity (left) and the power in rAI (right) time-locked to the attention switch event, which is represented by the dashed black trace. The y-axes denote all the trials of all subjects, and the intensity of the colors represent the respective magnitude of the EC and power time series. Next, in Fig. 5.5 we time-locked the connectivity and power data to the maximum of the rAI→ACC connectivity time series. Lastly, in Fig. 5.6 we plot the group-level trial average of the rAI→ACC connectivity time series (top) and rAI and ACC power (bottom), and the inset equation models empirically the activity of the ACC as a function of the activity in rAI and their coupling.

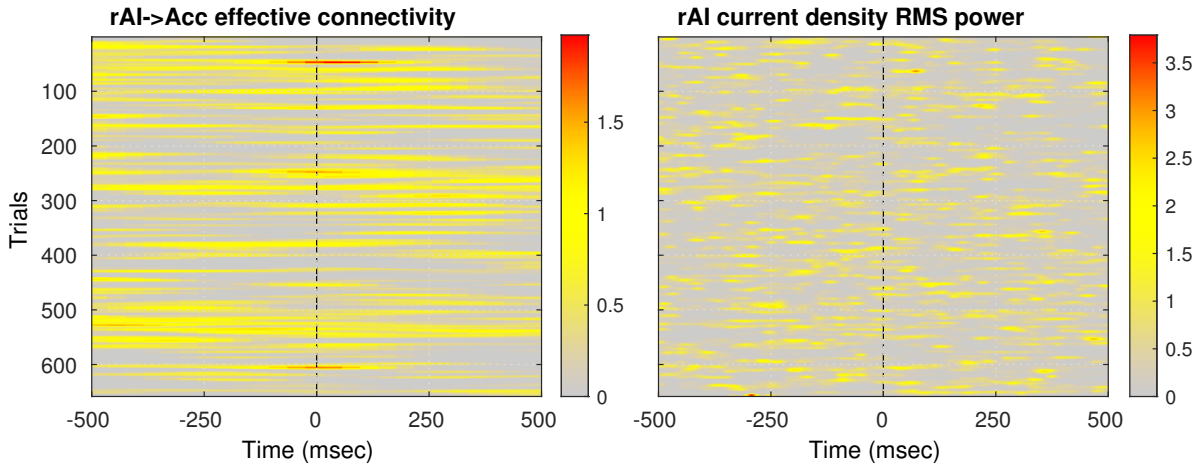


Figure 5.4. SN connectivity (left) and power (right) dynamics time-locked to the putative attention switch event. The y-axes denote all the trials of all subjects and the intensity of the colors represent the respective magnitude of the rAI→ACC connectivity and rAI power time series.

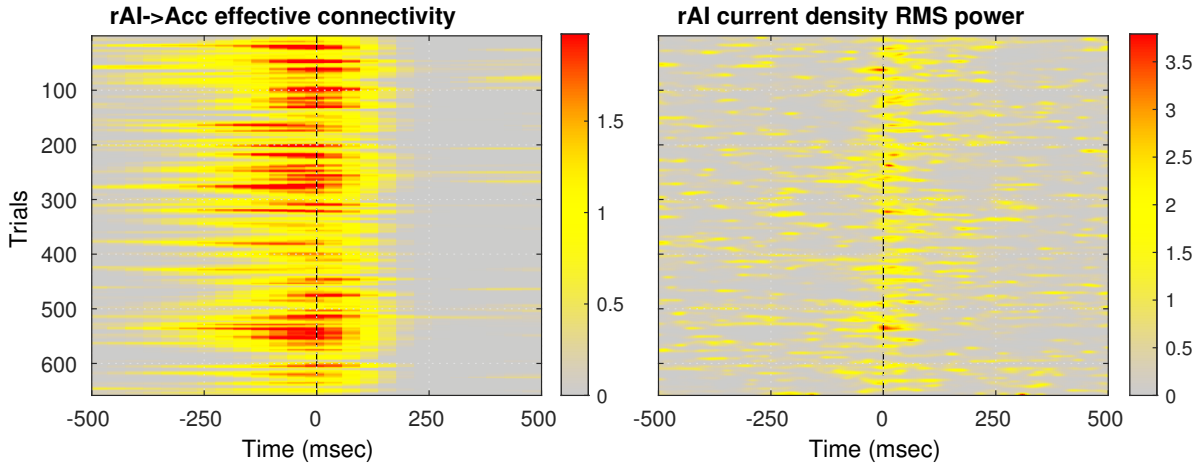


Figure 5.5. SN connectivity (left) and power (right) dynamics time-locked to the maximum of rAI→ACC connectivity time series. The y-axes denote all the trials of all subjects and the intensity of the colors represent the respective magnitude of rAI→ACC connectivity and rAI power time series.

5.4 Discussion and conclusions

The results shown in Fig. 5.4 indicate that there is no consistent activity in the SN time-locked to the attention switch event. This is not surprising because, given the unconstrained nature of our task, the last behavioral event in the on-task period does

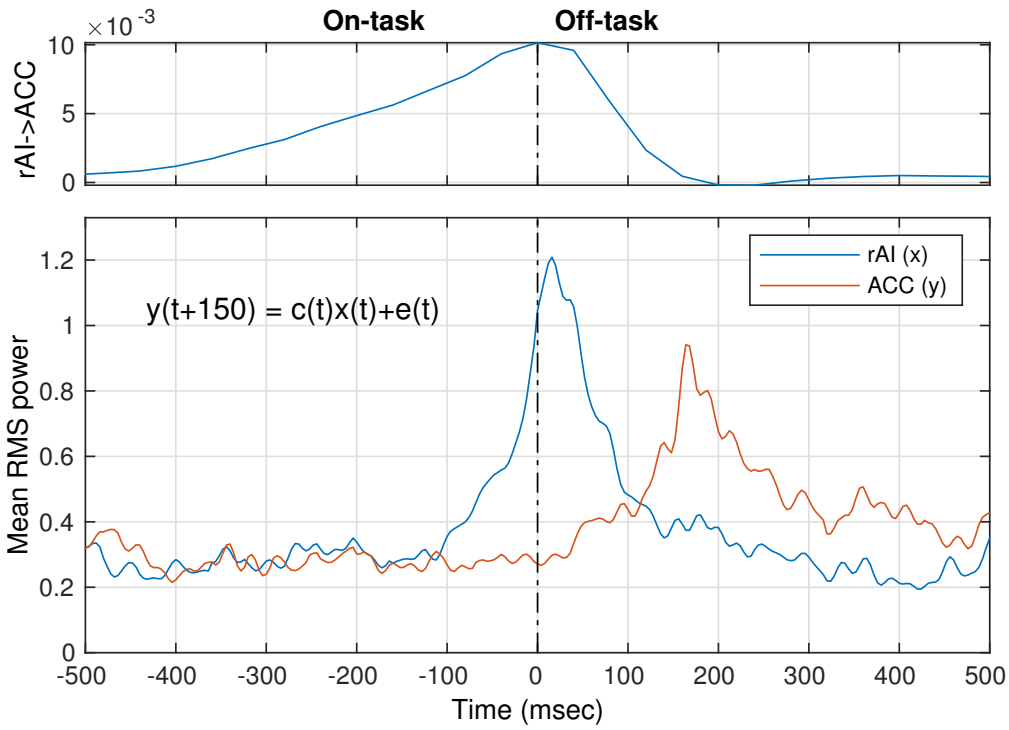


Figure 5.6. Group-level trial average of the rAI→ACC connectivity time series (top) and rAI and ACC power (botom).

not necessarily have to align with the deployment of attentional resources but with the fulfillment of a motor command. Fig. 5.5, however, tells a more interesting story. In that figure, we see that when we time-lock connectivity and power data to the maximum of the rAI→ACC connectivity there is a clear "internally-driven" evoked activity in the SN. In the left panel of that figure, we see that from left to right, the causal coupling rAI→ACC increases, reaches a peak, and then falls to a background value from which it does not recover. In the right panel, we also see a consistent increase in rAI power right after the connectivity reaches its peak. The results of Fig. 5.5 indicate that information flows from rAI to ACC in the vicinity of the attention switch event, and that this flow was consistent across trials and subjects, which is the hypothesis what we wanted to test at the outset. This interpretation is supported more clearly in Fig. 5.6, where we see that the peak of power activity in the rAI reaches the ACC approximately 150 msec after, thus indicating that these two regions are effectively (causally) connected.

We need more experiments to determine whether, in our unconstrained task, the observed SN activity is triggering a switch in attention, and thereby causing the observed behavioral change. At this point, our analysis can establish that there is a correlation between SN activity and behavior, but we cannot prove causation. We will further this analysis in future research to find a mechanistic explanation for internally-driven switches of attention. Taken as a whole, the results obtained in this chapter underscore the importance of the brain imaging methods developed in this dissertation for: 1) the discovery of brain mechanisms that support cognition and behavior, and 2) designing circuit-specific personalized BCI neurotherapies based on those mechanisms. For instance, if indeed rAI→ACC network activity causes our attention to switch, an imaging-based BCI could monitor the rAI→ACC coupling and the power in rAI, and when both increase significantly over a threshold, the BCI would know that an attention switch will be triggered ∼150 after.

5.5 Acknowledgements

This research was supported by NIMH Training Fellowship in Cognitive Neuroscience (AO), UC San Diego Chancellor's Research Excellence Scholarship (JM, AO), and UC San Diego School of Medicine start-up funds (JM).

This chapter is currently being prepared for submission for publication of the material. I am the primary investigator and author of this material, Margot Wagner and Jyoti Mishra are the second and senior authors.

Chapter 6

Conclusion

In this dissertation, I focused on the development of neuroimaging technology based on EEG that allows us to monitor the electrical activity of cortical networks in a noninvasive way, with low-cost, and high spatiotemporal resolution. This technology can serve as the "neural decoder" component of yet to come imaging-based closed-loop brain-computer interface (BCI) systems in which we can dynamically probe specific cognitive abilities in search for signatures of circuit dysfunctions while giving feedback targeted to inducing the engagement of neural populations that can compensate the aberrant neuronal activity. The motivation for such systems is twofold: 1) a neural decoder based on EEG source network activity can shed light on the neurobiological dysfunctions that lead to mental illnesses and 2) the use of features with biological relevance can facilitate the mapping from neural states to behavior.

To realize my vision of imaging-based neurotherapeutic BCIs, first in Chapter 2 I used biological and mathematical constraints to develop a coherent Bayesian framework (PEB+, Parametric Empirical Bayes with the addition of artifacts modeling) that allowed the unification of three of the most common problem in EEG analysis: data cleaning, source separation, and imaging. Chapters 3 and 4 I developed the algorithmic and software infrastructure to develop imaging-based BCIs. In particular, in Chapter 3 I redesigned an existent block-sparse Bayesian learning algorithm to yield efficient noise adaptation and

real-time performance. This algorithm is then used in the SimBSI library, introduced in Chapter 4 as the solver for the PEB+ framework. Finally, in Chapter 5, I applied these imaging methods to the study of the cortical networks involved in the switch of attention of healthy adult subjects during an unconstrained schoolwork-like computerized task. The results of this last chapter underscore the importance of using EEG source imaging with a high spatiotemporal resolution for studying the neurobiological mechanisms by which brain activity give rise to high-level cognition and organizes behavior.

As I look into the future, there are several directions in which this research can be extended. First, the artifact dictionary used in the PEB+ framework may be augmented over time to cover a wider variety of artifact sources. Also, although in PEB+ there are no regularization parameters to set because we learn them from data, we still need to select the size of the block over which we assume iid data samples so that we robustly estimate hyperparameters. To eliminate the iid assumption, future work could reformulate PEB+ into the state-space framework by incorporating temporal priors, thereby switching from batch updates to a proper sample-based spatiotemporal filter. On the data analysis side, we can further the study of the attention switch data to unveil the circuits that intervene in the regulation of attentional resources and the utilization of this knowledge in the development of closed-loop therapies that can increase the attention span of people with attention deficit disorders.

Appendix A

Equivalence between the the first stage of BSBL-2S and the evidence framework

The log evidence of a linear Gaussian model subject to the constraint of Eq. (3.2) is typically expressed as follows [Bishop, 2006][1],

$$\begin{aligned}\log p(\mathbf{Y}|\lambda,\gamma_F) = &-\frac{1}{2}\log|\gamma_F\mathbf{C}| - \frac{N_y}{2}\log\lambda - E(\hat{\mathbf{G}}) - \\ &\frac{1}{2}\log|\mathbf{A}| - \frac{N_y}{2}\log 2\pi\end{aligned} \tag{A.1}$$

where we have defined the following quantities,

$$E(\hat{\mathbf{g}}) = \lambda^{-1}\|\mathbf{Y} - \mathbf{L}\hat{\mathbf{G}}\|_F^2 + \|\hat{\mathbf{G}}\|_{(\gamma_F\mathbf{C})^{-1}}^2, \tag{A.2a}$$

$$\mathbf{A} = (\gamma_F\mathbf{C})^{-1} + \lambda^{-1}\mathbf{L}^T\mathbf{L} \tag{A.2b}$$

and $\hat{\mathbf{G}}$ is given by Eq. (2.10) evaluated in $\mathbf{Y}$. We reduce Eq. (C.2a) using the following identity

$$\text{trace}(\mathbf{C}_y\boldsymbol{\Sigma}_y^{-1}) = \lambda^{-1}\|\mathbf{Y} - \mathbf{L}\hat{\mathbf{g}}\|_F^2 + \|\hat{\mathbf{G}}\|_{(\gamma_F\mathbf{C})^{-1}}^2 \tag{A.3}$$

[1] In that reference $\mathbf{C} = \mathbf{I}_{N_g}$.

Next, we reduce Eq. (C.2b) using the matrix determinant lemma,

$$\log|\mathbf{A}| = \log|\mathbf{\Sigma}_y| - N_y \log\lambda - \log|\gamma_F \mathbf{C}| \tag{A.4}$$

Substituting Eq. (C.3) and Eq. (C.4) in Eq. (C.1), dropping quantities that do not depend on λ or γ_F, and multiplying by -2 we obtain $\mathcal{L}(\mathcal{M}_F)$,

$$-2\log p(\mathbf{Y}|\lambda, \gamma_F) = \log|\mathbf{\Sigma}_y| + \text{trace}(\mathbf{C}_y \mathbf{\Sigma}_y^{-1}) \tag{A.5}$$

Furthermore, using the *SVD* of $\mathbf{L}\Delta^{-1}$ we simplify the following expressions,

$$\begin{aligned}
\mathbf{\Sigma}_y &= \lambda \mathbf{I}_{N_y} + \gamma_F \mathbf{U}\text{diag}(s_i^2)\mathbf{U}^T \\
&= \mathbf{U}\text{diag}(\psi_i)\mathbf{U}^T \\
\mathbf{\Sigma}_y^{-1} &= \mathbf{U}\text{diag}(\psi_i^{-1})\mathbf{U}^T \\
\log|\mathbf{\Sigma}_y| &= \sum_i^{N_y} \log\psi_i
\end{aligned} \tag{A.6}$$

Finally, inserting Eq. (C.6) in Eq. (C.5) we write the ML-II functional as a function of the regularized eigenvalues ψ_i in Eq. (3.8).

Appendix B

Algebraic manipulations to avoid computing the inverse of the source prior covariance

In Chapter 3, we defined the known ("designed") source prior covariance to have a block-diagonal structure. This block-diagonal matrix can be defined in terms of a Laplacian operator constructed from the cortical mesh, $\mathbf{C} = (\Delta^T \Delta)^{-1}$. Note that the matrix Δ is the matrix square root of C. The matrix Δ is sparse with the following block structure:

$$\Delta = \begin{bmatrix} \Delta_1 & & \\ & \ddots & \\ & & \Delta_n \end{bmatrix} \tag{B.1}$$

where each Δ_i cell corresponds to a Laplacian operator defined on the ith group only. From linear algebra we have that its inverse can be computed as the inverse of each block as follows:

$$\Delta^{-1} = \begin{bmatrix} \Delta_1^{-1} & & \\ & \ddots & \\ & & \Delta_n^{-1} \end{bmatrix} \tag{B.2}$$

Then, **abusing the notation**, let's define the ith square root covariance Δ_i^{-1} by zeroing out the entries in Eq. (B.2) outside the ith group such that:

$$\Delta_i^{-1} \triangleq \begin{bmatrix} \mathbf{0}_1 & & & \\ & \ddots & & \\ & & \Delta_i^{-1} & \\ & & & \ddots \\ & & & & \mathbf{0}_n \end{bmatrix} \tag{B.3}$$

Note that we compute the inverse before applying the ith mask, otherwise the ith inverse won't exist. Finally we obtain the standardized observation operators as:

$$\mathbf{L}_i = \mathbf{L}\Delta_i^{-1} \tag{B.4}$$

In practice, to reduced even moore the computational footprint of Eq. (B.4), we can select the columns in $\mathbf{L}$ that belong to the ith group and multiply that submatrix with the Δ_i^{-1} cell that appears in the right hand side of Eq. (B.2).

Appendix C

Fast two-stage BSBL algorithm

The log evidence of a linear Gaussian model subject to the constraint of Eq. (3.2) is typically expressed as follows [Bishop, 2006][1],

$$\begin{aligned}\log p(\mathbf{Y}|\lambda,\gamma_F) = &-\frac{1}{2}\log|\gamma_F\mathbf{C}| - \frac{N_y}{2}\log\lambda - E(\hat{\mathbf{G}}) - \\ &\frac{1}{2}\log|\mathbf{A}| - \frac{N_y}{2}\log 2\pi\end{aligned} \tag{C.1}$$

where we have defined the following quantities,

$$E(\hat{\mathbf{g}}) = \lambda^{-1}\|\mathbf{Y} - \mathbf{L}\hat{\mathbf{G}}\|_F^2 + \|\hat{\mathbf{G}}\|_{(\gamma_F\mathbf{C})^{-1}}^2, \tag{C.2a}$$

$$\mathbf{A} = (\gamma_F\mathbf{C})^{-1} + \lambda^{-1}\mathbf{L}^T\mathbf{L} \tag{C.2b}$$

and $\hat{\mathbf{G}}$ is given by Eq. (2.10) evaluated in $\mathbf{Y}$. We reduce Eq. (C.2a) using the following identity

$$\text{trace}(\mathbf{C}_y\boldsymbol{\Sigma}_y^{-1}) = \lambda^{-1}\|\mathbf{Y} - \mathbf{L}\hat{\mathbf{g}}\|_F^2 + \|\hat{\mathbf{G}}\|_{(\gamma_F\mathbf{C})^{-1}}^2 \tag{C.3}$$

Next, we reduce Eq. (C.2b) using the matrix determinant lemma,

$$\log|\mathbf{A}| = \log|\boldsymbol{\Sigma}_y| - N_y\log\lambda - \log|\gamma_F\mathbf{C}| \tag{C.4}$$

[1] In that reference $\mathbf{C} = \mathbf{I}_{N_g}$.

Substituting Eq. (C.3) and Eq. (C.4) in Eq. (C.1), dropping quantities that do not depend on λ or γ_F, and multiplying by -2 we obtain $\mathcal{L}(\mathcal{M}_F)$,

$$-2\log p(\mathbf{Y}|\lambda,\gamma_F) = \log|\boldsymbol{\Sigma}_y| + \mathrm{trace}(\mathbf{C}_y\boldsymbol{\Sigma}_y^{-1}) \tag{C.5}$$

Furthermore, using the *SVD* of $\mathbf{L}\Delta^{-1}$ we simplify the following expressions,

$$\begin{aligned}
\boldsymbol{\Sigma}_y =& \lambda\mathbf{I}_{N_y} + \gamma_F\mathbf{U}\mathrm{diag}(s_i^2)\mathbf{U}^T \\
=& \mathbf{U}\mathrm{diag}(\psi_i)\mathbf{U}^T \\
\boldsymbol{\Sigma}_y^{-1} =& \mathbf{U}\mathrm{diag}(\psi_i^{-1})\mathbf{U}^T \\
\log|\boldsymbol{\Sigma}_y| =& \sum_i^{N_y}\log\psi_i
\end{aligned} \tag{C.6}$$

Finally, inserting Eq. (C.6) in Eq. (C.5) we write the ML-II functional as a function of the regularized eigenvalues ψ_i in Eq. (3.8).

Bibliography

Akil, H., Brenner, S., Kandel, E., Kendler, K. S., King, M. C., Scolnick, E., Watson, J. D., and Zoghbi, H. Y. (2010). The future of psychiatric research: Genomes and neural circuits.

Akinbami, L. J., Liu, X., Pastor, P. N., and Reuben, C. A. (2011). Attention deficit hyperactivity disorder among children aged 5-17 years in the United States, 1998-2009. *NCHS data brief*.

Arnsten, A. F. T. and Rubia, K. (2012). Neurobiological circuits regulating attention, cognitive control, motivation, and emotion: Disruptions in neurodevelopmental psychiatric disorders.

Artoni, F., Delorme, A., and Makeig, S. (2018). Applying dimension reduction to EEG data by Principal Component Analysis reduces the quality of its subsequent Independent Component decomposition. *NeuroImage*.

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

Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge University Press.

Bateson, A. D., Baseler, H. A., Paulson, K. S., Ahmed, F., and Asghar, A. U. R. (2017). Categorisation of Mobile EEG: A Researcher's Perspective. *BioMed Research International*, 2017:1–15.

Bauer, F. and Lukas, M. A. (2011). Comparing parameter choice methods for regularization of ill-posed problems. *Mathematics and Computers in Simulation*, 81(9):1795–1841.

Bauer, H. and Pllana, A. (2014). EEG-based local brain activity feedback training-tomographic neurofeedback. *Frontiers in Human Neuroscience*, 8.

Baumann, O. and Mattingley, J. B. (2010). Medial Parietal Cortex Encodes Perceived Heading Direction in Humans. *Journal of Neuroscience*, 30(39):12897–901.

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

Bertero, M., De Mol, C., and Pike, E. R. (1988). Linear inverse problems with discrete data : 11 . Stability and regularisation. *Inverse Problems*, 4:573–594.

Bigdely-Shamlo, N., Kreutz-Delgado, K., Kothe, C., and Makeig, S. (2013a). EyeCatch: Data-mining over half a million EEG independent components to construct a fully-automated eye-component detector. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5845–5848. IEEE.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., and Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*.

Bigdely-Shamlo, N., Mullen, T., Kreutz-Delgado, K., and Makeig, S. (2013b). Measure projection analysis: A probabilistic approach to EEG source comparison and multi-subject inference. *NeuroImage*.

Bishop, C. M. C. C. M. (2006). Pattern Recognition and Machine Learning. *Pattern Recognition*, 4(4):738.

Böl, M., Weikert, R., and Weichert, C. (2011). A coupled electromechanical model for the excitation-dependent contraction of skeletal muscle. *Journal of the Mechanical Behavior of Biomedical Materials*.

Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122.

Breakspear, M. (2017). Dynamic models of large-scale brain activity. *Nature neuroscience*.

Brunner, C., Andreoni, G., Bianchi, L., Blankertz, B., Breitwieser, C., Kanoh, S., Kothe, C. A., Lécuyer, A., Makeig, S., Mellinger, J., Perego, P., Renard, Y., Schalk, G., Susila, I. P., Venthur, B., and Müller-Putz, G. R. (2012). BCI Software Platforms. In Allison, B. Z., Dunne, S., Leeb, R., Del R. Millán, J., and Nijholt, A., editors, *Towards Practical Brain-Computer Interfaces*, Biological and Medical Physics, Biomedical Engineering, pages 303–331. Springer Berlin Heidelberg, Berlin, Heidelberg.

Brunner, C., Birbaumer, N., Blankertz, B., Guger, C., Kübler, A., Mattia, D., Millán, J. d. R., Miralles, F., Nijholt, A., Opisso, E., Ramsey, N., Salomon, P., and Müller-Putz, G. R. (2015). BNCI Horizon 2020: towards a roadmap for the BCI community. *Brain-Computer Interfaces*, 2(1):1–10.

Burke, J. F., Merkow, M. B., Jacobs, J., Kahana, M. J., and Zaghloul, K. A. (2015). Brain computer interface to enhance episodic memory in human participants. *Frontiers in Human Neuroscience*.

Buzzell, G. A., Richards, J. E., White, L. K., Barker, T. V., Pine, D. S., and Fox, N. A.

(2017). Development of the error-monitoring system from ages 9âĂŞ35: Unique insight provided by MRI-constrained source localization of EEG. *NeuroImage*, 157:13–26.

Byrne, P., Becker, S., and Burgess, N. (2007). Remembering the past and imagining the future: A neural model of spatial memory and imagery. *Psychological Review*.

Carelli, L., Solca, F., Faini, A., Meriggi, P., Sangalli, D., Cipresso, P., Riva, G., Ticozzi, N., Ciammola, A., Silani, V., and Poletti, B. (2017). Brain-Computer Interface for Clinical Purposes: Cognitive Assessment and Rehabilitation.

Casella, G. (1985). An introduction to empirical Bayes data analysis. *The American Statistician*, 39(2):83–87.

Chang, C.-Y., Hsu, S.-H., Pion-Tonachini, L., and Jung, T.-P. (2018). Evaluation of Artifact Subspace Reconstruction for Automatic EEG Artifact Removal. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1242–1245. IEEE.

Chavarriaga, R. and del R. Millán, J. (2010). Learning From EEG Error-Related Potentials in Noninvasive Brain-Computer Interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(4):381–388.

Chavarriaga, R., Sobolewski, A., and Millan, J. d. R. (2014). Errare machinale est: The use of error-related potentials in brain-machine interfaces.

Chiesi, M., Guermandi, M., Placati, S., Franchi Scarselli, E., and Guerrieri, R. (2018). Creamino: a Cost-Effective, Open-Source EEG-based BCI System. *IEEE Transactions on Biomedical Engineering*, pages 1–1.

Chiu, T. C., Gramann, K., Ko, L. W., Duann, J. R., Jung, T. P., and Lin, C. T. (2012). Alpha modulation in parietal and retrosplenial cortex correlates with navigation performance. *Psychophysiology*.

Chowdhury, R. A., Lina, J. M., Kobayashi, E., and Grova, C. (2013). MEG Source Localization of Spatially Extended Generators of Epileptic Activity: Comparing Entropic and Hierarchical Bayesian Approaches. *PLoS ONE*, 8(2):e55969.

Cichocki, A. and Amari, S.-i. (2002). *Adaptive Blind Signal and Image Processing*. John Wiley & Sons, Ltd, Chichester, UK.

Combettes, P. L. and Pesquet, J.-C. (2011). Proximal Splitting Methods in Signal Processing. In *Springer Optimization and Its Applications*, volume 49, pages 185–212.

Comon, P. (1994). Independent component analysis, A new concept? *Signal Processing*,

36(3):287–314.

Congedo, M., Barachant, A., and Bhatia, R. (2017). Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces*.

Conners, K. K., Epstein, J. N., Angold, A., and Klaric, J. (2003). Continuous performance test performance in a normative epidemiological sample. *Journal of Abnormal Child Psychology*.

Costa, Ã., Salazar-Varas, R., Úbeda, A., and Azorín, J. M. (2016). Characterization of Artifacts Produced by Gel Displacement on Non-invasive Brain-Machine Interfaces during Ambulation. *Frontiers in Neuroscience*, 10.

Cotter, S., Rao, B. D., and Kreutz-Delgado, K. (2005). Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Transactions on Signal Processing*, 53(7):2477–2488.

Courellis, H., Mullen, T., Poizner, H., Cauwenberghs, G., and Iversen, J. R. (2017). EEG-Based Quantification of Cortical Current Density and Dynamic Causal Connectivity Generalized across Subjects Performing BCI-Monitored Cognitive Tasks. *Frontiers in Neuroscience*, 11.

CSM (2015). The Common Sense Census: Media Use by Tweens and Teens | Common Sense Media. Technical report.

Cuspineda, E., Machado, C., Virues, T., Martínez-Montes, E., Ojeda, A., Valdés, P. A., Bosch, J., and Valdes, L. (2009). Source Analysis of Alpha Rhythm Reactivity Using LORETA Imaging with 64-Channel EEG and Individual MRI. *Clinical EEG and Neuroscience*, 40(3):150–156.

Dale, A. M. and Sereno, M. I. (1993). Improved Localizadon of Cortical Activity by Combining EEG and MEG with MRI Cortical Surface Reconstruction: A Linear Approach. *Journal of Cognitive Neuroscience*, 5(2):162–176.

David, O., Kiebel, S. J., Harrison, L. M., Mattout, J., Kilner, J. M., and Friston, K. J. (2006). Dynamic causal modeling of evoked responses in EEG and MEG. *NeuroImage*, 30(4):1255–1272.

De Martino, E., Seminowicz, D. A., Schabrun, S. M., Petrini, L., and Graven-Nielsen, T. (2019). High frequency repetitive transcranial magnetic stimulation to the left dorsolateral prefrontal cortex modulates sensorimotor cortex function in the transition to sustained muscle pain. *NeuroImage*, 186:93–102.

Delorme, A., Miyakoshi, M., Jung, T. P., and Makeig, S. (2015). Grand average ERP-image

plotting and statistics: A method for comparing variability in event-related single-trial EEG activities across subjects and conditions. *Journal of Neuroscience Methods*, 250:3–6.

Delorme, A., Mullen, T., Kothe, C., Akalin Acar, Z., Bigdely-Shamlo, N., Vankov, A., and Makeig, S. (2011). EEGLAB, SIFT, NFT, BCILAB, and ERICA: New tools for advanced EEG processing. *Computational Intelligence and Neuroscience*, 2011.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., and Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS ONE*.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., and Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980.

Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*.

Di Martino, A., Ghaffari, M., Curchack, J., Reiss, P., Hyde, C., Vannucci, M., Petkova, E., Klein, D. F., and Castellanos, F. X. (2008). Decomposing Intra-Subject Variability in Children with Attention-Deficit/Hyperactivity Disorder. *Biological Psychiatry*.

Drajsajtl, T., Struk, P., and Bednárová, A. (2013). AsTeRICS. In *Studies in Health Technology and Informatics*.

Duque-Munoz, L., Vargas, F., and Lopez, J. D. (2016). Simplified EEG inverse solution for BCI real-time implementation. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, volume 2016-Octob, pages 4051–4054.

Edelman, B. J., Baxter, B., and He, B. (2016). EEG source imaging enhances the decoding of complex right-hand motor imagery tasks. *IEEE Transactions on Biomedical Engineering*, 63(1):4–14.

Efron, B. and Morris, C. (1972a). Limiting the Risk of Bayes and Empirical Bayes Estimators–Part I: The Bayes Case. *Journal of the American Statistical Association*, 67(337):130.

Efron, B. and Morris, C. (1972b). Limiting the Risk of Bayes and Empirical Bayes Estimators–Part II: The Empirical Bayes Case. *Journal of the American Statistical Association*, 67(337):130.

Esterman, M., Noonan, S. K., Rosenberg, M., and Degutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention.

Cerebral Cortex.

Fortenbaugh, F. C., Degutis, J., and Esterman, M. (2017). Recent theoretical, neural, and clinical advances in sustained attention research.

Fortenbaugh, F. C., Degutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., and Esterman, M. (2015). Sustained attention across the life span in a sample of 10,000: Dissociating ability and strategy. *Psychological Science*.

Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., and Mattout, J. (2008). Multiple sparse priors for the M/EEG inverse problem. *NeuroImage*, 39(3):1104–1120.

Friston, K. and Penny, W. (2011). Post hoc Bayesian model selection. *NeuroImage*, 56(4):2089–2099.

Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J., Glaser, D. E., Henson, R. N. a., Kiebel, S., Phillips, C., and Ashburner, J. (2002). Classical and Bayesian Inference in Neuroimaging: Theory. *NeuroImage*, 16(2):465–483.

Fu, M. J., Daly, J. J., and Cavusoglu, M. C. (2006). A Detection Scheme for Frontalis and Temporalis Muscle EMG Contamination of EEG Data. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4514–4518. IEEE.

Fujiwara, Y., Yamashita, O., Kawawaki, D., Doya, K., Kawato, M., Toyama, K., and Sato, M.-a. (2009). A hierarchical Bayesian method to resolve an inverse problem of MEG contaminated with eye movement artifacts. *NeuroImage*, 45(2):393–409.

Gazzaley, A. and Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory.

Gehring, W. J., Liu, Y., Orr, J. M., and Carp, J. (2012). The Error-Related Negativity (ERN/Ne). In *The Oxford Handbook of Event-Related Potential Components*.

Georgieva, P., Silva, F., Milanova, M., and Kasabov, N. (2014). EEG Signal Processing for BrainâĂŞComputer Interfaces. In *Springer Handbook of Bio-/Neuroinformatics*, pages 797–812. Springer Berlin Heidelberg, Berlin, Heidelberg.

Giraldo, E., den Dekker, a. J., and Castellanos-Dominguez, G. (2010). Estimation of dynamic neural activity using a Kalman filter approach based on physiological models. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2010:2914–7.

Golub, G. H., Heath, M., and Wahba, G. (1979). Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics*, 21(2):215–223.

Gorodnitsky, I. F. and Rao, B. D. (1997). Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, 45(3):600–616.

Gramann, K., Ferris, D. P., Gwin, J., and Makeig, S. (2014). Imaging natural cognition in action. *International Journal of Psychophysiology*, 91(1):22–29.

Gramann, K., Hohlefeld, F. U., Gehrke, L., and Klug, M. (2018). Heading computation in the human retrosplenial complex during full-body rotation. *bioRxiv*.

Gramann, K., Onton, J., Riccobon, D., Mueller, H. J., Bardins, S., and Makeig, S. (2010). Human brain dynamics accompanying use of egocentric and allocentric reference frames during navigation. *Journal of Cognitive Neuroscience*.

Gramfort, A., Papadopoulo, T., Olivi, E., and Clerc, M. (2010). OpenMEEG: opensource software for quasistatic bioelectromagnetics. *Biomedical engineering online*, 9:45.

Gramfort, a., Strohmeier, D., Haueisen, J., Hämäläinen, M. S., and Kowalski, M. (2013). Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations. *NeuroImage*, 70:410–22.

Guger, C., Schlögl, A., Neuper, C., Walterspacher, D., Strain, T., and Pfurtscheller, G. (2001). Rapid prototyping of an EEG-based brain-computer interface (BCI). *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.

Hallez, H., Vanrumste, B., Grech, R., Muscat, J., De Clercq, W., Vergult, A., D'Asseler, Y., Camilleri, K. P., Fabri, S. G., Van Huffel, S., and Lemahieu, I. (2007). Review on solving the forward problem in EEG source analysis. *Journal of NeuroEngineering and Rehabilitation*, 4(1):46.

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(2):851–859.

Henson, R. N., Wakeman, D. G., Litvak, V., and Friston, K. J. (2011). A Parametric Empirical Bayesian Framework for the EEG/MEG Inverse Problem: Generative Models for Multi-Subject and Multi-Modal Integration. *Frontiers in Human Neuroscience*, 5.

Herrmann, M. J., Römmler, J., Ehlis, A.-C., Heidrich, A., and Fallgatter, A. J. (2004). Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). *Cognitive Brain Research*, 20(2):294–299.

Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., and Evans, A. C. (1998). Enhancement of MR Images Using Registration for Signal Averaging. *Journal of Computer Assisted Tomography*, 22(2):324–333.

Hsu, S.-H., Mullen, T. R., Jung, T.-P., and Cauwenberghs, G. (2016). Real-Time Adaptive EEG Source Separation Using Online Recursive Independent Component Analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.

Huang, M. X., Dale, A. M., Song, T., Halgren, E., Harrington, D. L., Podgorny, I., Canive, J. M., Lewis, S., and Lee, R. R. (2006). Vector-based spatial-temporal minimum L1-norm solution for MEG. *NeuroImage*.

Huang, Y., Parra, L. C., and Haufe, S. (2016). The New York HeadâĂŤA precise standardized volume conductor model for EEG source localization and tES targeting. *NeuroImage*.

Hurtier, J., Van Dokkum, L., Dalhoumi, S., Coffey, A., Perrey, S., Jourdan, C., Dray, G., Ward, T., Froger, J., and Laffont, I. (2016). A closed-loop BCI system for rehabilitation of the hemiplegic upper-limb: A performance study of the systems ability to detect intention of movement. *Annals of Physical and Rehabilitation Medicine*.

Islam, M. K., Rastegarnia, A., and Yang, Z. (2016). Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiologie Clinique/Clinical Neurophysiology*.

Janani, A. S., Grummett, T. S., Lewis, T. W., Fitzgibbon, S. P., Whitham, E. M., DelosAngeles, D., Bakhshayesh, H., Willoughby, J. O., and Pope, K. J. (2017). Evaluation of a minimum-norm based beamforming technique, sLORETA, for reducing tonic muscle contamination of EEG at sensor level. *Journal of Neuroscience Methods*, 288:17–28.

Jeunet, C., Jahanpour, E., and Lotte, F. (2016). Why standard brain-computer interface (BCI) training protocols should be changed: An experimental study. *Journal of Neural Engineering*, 13(3).

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., and Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*.

Jungnickel, E. and Gramann, K. (2016). Mobile Brain/Body Imaging (MoBI) of Physical Interaction with Dynamically Moving Objects. *Frontiers in Human Neuroscience*, 10.

Karabanov, A., Thielscher, A., and Siebner, H. R. (2016). Transcranial brain stimulation: Closing the loop between brain and stimulation.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical*

Association, 90(430):773–795.

Khambhati, A. N., Sizemore, A. E., Betzel, R. F., and Bassett, D. S. (2018). Modeling and interpreting mesoscale network dynamics. *NeuroImage*, 180:337–349.

Khanna, A., Pascual-Leone, A., Michel, C. M., and Farzan, F. (2015). Microstates in resting-state EEG: Current status and future directions. *Neuroscience & Biobehavioral Reviews*, 49:105–113.

Kiebel, S. J., Garrido, M. I., Moran, R., Chen, C.-C., and Friston, K. J. (2009). Dynamic causal modeling for EEG and MEG. *Human brain mapping*, 30(6):1866–1876.

Kilicarslan, A., Grossman, R. G., and Contreras-Vidal, J. L. (2016). A robust adaptive denoising framework for real-time artifact removal in scalp EEG measurements. *Journal of Neural Engineering*, 13(2):026013.

Koenig, T., Prichep, L., Lehmann, D., Sosa, P. V., Braeker, E., Kleinlogel, H., Isenhart, R., and John, E. (2002). Millisecond by Millisecond, Year by Year: Normative EEG Microstates and Developmental Stages. *NeuroImage*, 16(1):41–48.

Kothe, C. A. and Makeig, S. (2013). BCILAB: A platform for brain-computer interface development. *Journal of Neural Engineering*.

Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Physical Review E*, 69(6):066138.

Kumari, P., Mathew, L., and Syal, P. (2017). Increasing trend of wearables and multimodal interface for human activity monitoring: A review. *Biosensors and Bioelectronics*, 90:298–307.

Le, Q. V., Karpenko, A., Ngiam, J., and Ng, A. (2011). ICA with reconstruction cost for efficient overcomplete feature learning. In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K., editors, *Advances in Neural Information Processing Systems 24*, pages 1017–1025. Curran Associates, Inc.

Lee, T.-S., Goh, S. J. A., Quek, S. Y., Phillips, R., Guan, C., Cheung, Y. B., Feng, L., Teng, S. S. W., Wang, C. C., Chin, Z. Y., Zhang, H., Ng, T. P., Lee, J., Keefe, R., and Krishnan, K. R. R. (2013). A Brain-Computer Interface Based Cognitive Training System for Healthy Elderly: A Randomized Control Pilot Study for Usability and Preliminary Efficacy. *PLoS ONE*, 8(11):e79419.

Lei, X., Wu, T., and Valdes-Sosa, P. A. (2015). Incorporating priors for EEG source imaging and connectivity analysis. *Frontiers in neuroscience*, 9(August):284.

Lewicki, M. S. and Sejnowski, T. J. (1998). Learning nonlinear overcomplete representations for efficient coding. *Advances in Neural Information Processing Systems 10 (NIPS'97)*, pages 556–562.

Liang, S. F., Shaw, F. Z., Young, C. P., Chang, D. W., and Liao, Y. C. (2010). A closed-loop brain computer interface for real-time seizure detection and control. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10*.

Lin, C. T., Chiu, T. C., and Gramann, K. (2015). EEG correlates of spatial orientation in the human retrosplenial complex. *NeuroImage*.

Lin, C. T., Chiu, T. C., Wang, Y. K., Chuang, C. H., and Gramann, K. (2018). Granger causal connectivity dissociates navigation networks that subserve allocentric and egocentric path integration. *Brain Research*.

Lizier, J. T. (2014). JIDT: An information-theoretic toolkit for studying the dynamics of complex systems.

Lizier, J. T., Prokopenko, M., and Zomaya, A. Y. (2008). Local information transfer as a spatiotemporal filter for complex systems. *Physical Review E*, 77(2):026110.

Lopes da Silva, F. (2013). EEG and MEG: Relevance to Neuroscience. *Neuron*, 80(5):1112–1128.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., and Yger, F. (2018). A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update.

Lotte, F., Larrue, F., and Mühl, C. (2013). Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design. *Frontiers in Human Neuroscience*, 7.

Luu, P., Tucker, D. M., and Makeig, S. (2004). Frontal midline theta and the error-related negativity: Neurophysiological mechanisms of action regulation. *Clinical Neurophysiology*, 115(8):1821–1835.

MacKay, D. J. C. (1992). Bayesian Interpolation. *Neural Computation*, 4(3):415–447.

MacKay, D. J. C. (1996). Hyperparameters: Optimize, or integrate out? *Maximum entropy and bayesian methods*, pages 43–59.

MacKay, D. J. C. (2008a). Independent Component Analysis and Latent Variable Modelling. In *Information Theory, Inference, and Learning Algorithms*, pages 437–440. Cambridge

University Press.

MacKay, D. J. C. (2008b). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

Makeig, S., Gramann, K., Jung, T.-P. P., Sejnowski, T. J., and Poizner, H. (2009). Linking brain, mind and behavior. *International Journal of Psychophysiology*, 73(2):95–100.

Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., and Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences of the United States of America*.

Makeig, S. and Onton, J. (2011). *ERP Features and EEG Dynamics*. Oxford University Press.

Malter Cohen, M., Jing, D., Yang, R. R., Tottenham, N., Lee, F. S., and Casey, B. J. (2013). Early-life stress has persistent effects on amygdala function and development in mice and humans. *Proceedings of the National Academy of Sciences*.

Mannan, M. M. N., Kamran, M. A., and Jeong, M. Y. (2018). Identification and Removal of Physiological Artifacts From Electroencephalogram Signals: A Review. *IEEE Access*, 6:30630–30652.

Martínez-Vargas, J. D., Grisales-Franco, F. M., and Castellanos-Dominguez, G. (2015). Estimation of M/EEG Non-stationary Brain Activity Using Spatio-temporal Sparse Constraints. In *Artificial Computation in Biology and Medicine*, pages 429–438. Springer International Publishing.

MathWorks (2018a). Choosing a simulation mode in Simulink.

MathWorks (2018b). Stateflow.

Matiko, J. W., Beeby, S., and Tudor, J. (2013). Real time eye blink noise removal from EEG signals using morphological component analysis. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2013(2):13–6.

Mcdowell, K., Lin, C. T., Oie, K. S., Jung, T. P., Gordon, S., Whitaker, K. W., Li, S. Y., Lu, S. W., and Hairston, W. D. (2013). Real-world neuroimaging technologies. *IEEE Access*.

McFarland, D. J., Daly, J., Boulay, C., and Parvaz, M. A. (2017). Therapeutic applications of BCI technologies. *Brain-Computer Interfaces*, 4(1-2):37–52.

McLoughlin, G., Palmer, J. A., Rijsdijk, F., and Makeig, S. (2014). Genetic Overlap between Evoked Frontocentral Theta-Band Phase Variability, Reaction Time Variability, and Attention-Deficit/Hyperactivity Disorder Symptoms in a Twin Study. *Biological Psychiatry*, 75(3):238–247.

Mehta, R. K. and Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Frontiers in Human Neuroscience*.

Menon, V. and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, 214(5-6):655–667.

Michel, C. M. and Murray, M. M. (2012). Towards the utilization of EEG as a brain imaging tool. *NeuroImage*, 61(2):371–385.

Millan, M. J., Agid, Y., Brüne, M., Bullmore, E. T., Carter, C. S., Clayton, N. S., Connor, R., Davis, S., Deakin, B., Derubeis, R. J., Dubois, B., Geyer, M. A., Goodwin, G. M., Gorwood, P., Jay, T. M., Joëls, M., Mansuy, I. M., Meyer-Lindenberg, A., Murphy, D., Rolls, E., Saletu, B., Spedding, M., Sweeney, J., Whittington, M., and Young, L. J. (2012). Cognitive dysfunction in psychiatric disorders: Characteristics, causes and the quest for improved therapy.

Mishra, J., DeVillers-Sidani, E., Merzenich, M., and Gazzaley, A. (2014). Adaptive training diminishes distractibility in aging across species. *Neuron*.

Mishra, J. and Gazzaley, A. (2014). Closed-Loop Rehabilitation of Age-Related Cognitive Disorders. *Seminars in Neurology*, 34(05):584–590.

Mishra, J. and Gazzaley, A. (2016). Cross-species approaches to cognitive neuroplasticity research. *NeuroImage*.

Morris, C. N. (1983). Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association*, 78(381):47–55.

Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Jung, T.-P., and Cauwenberghs, G. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Transactions on Biomedical Engineering*, 62(11):2553–2567.

Müller-Putz, G. R. (2011). Tools for brain-computer interaction: a general concept for a hybrid BCI. *Frontiers in Neuroinformatics*.

Narayanan, N. S., Cavanagh, J. F., Frank, M. J., and Laubach, M. (2013). Common medial frontal mechanisms of adaptive control in humans and rodents. *Nature Neuroscience*.

Neal, R. M. (1996). *Bayesian Learning for Neural Networks*, volume 118 of *Lecture Notes*

in Statistics. Springer New York, New York, NY.

Noirhomme, Q., Kitney, R. I., and Macq, B. (2008). Single-trial EEG source reconstruction for brain-computer interface. *IEEE Transactions on Biomedical Engineering*, 55(5):1592–1601.

Nunez, P. L. and Srinivasan, R. (2006). *Electric Fields of the Brain*. Oxford University Press.

Ojeda, A., Bigdely-Shamlo, N., and Makeig, S. (2014). MoBILAB: an open source toolbox for analysis and visualization of mobile brain/body imaging data. *Frontiers in Human Neuroscience*, 8:121.

Ojeda, A., Klug, M., Kreutz-Delgado, K., Gramann, K., and Mishra, J. (2019). A Bayesian framework for unifying data cleaning, source separation and imaging of electroencephalographic signals.

Ojeda, A., Kreutz-Delgado, K., and Mullen, T. (2018). Fast and robust Block-Sparse Bayesian learning for EEG source imaging. *NeuroImage*, 174(8):449–462.

Onton, J., Westerfield, M., Townsend, J., and Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience & Biobehavioral Reviews*, 30(6):808–822.

Oostendorp, T. F., van Oosterom, A., and Huiskamp, G. (1989). Interpolation on a triangulated 3D surface. *Journal of Computational Physics*, 80(2):331–343.

Ophir, E., Nass, C., and Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences*.

Palmer, J., Kreutz-Delgado, K., and Makeig, S. (2011). AMICA: An Adaptive Mixture of Independent Component Analyzers with Shared Components. *San Diego, CA: Technical report, Swartz Center for Computational Neuroscience*.

Pascual-Leone, A., Rubio, B., Pallardó, F., and Catalá, M. D. (1996). Rapid-rate transcranial magnetic stimulation of left dorsolateral prefrontal cortex in drug-resistant depression. *The Lancet*, 348(9022):233–237.

Pascual-Marqui, R. D., Esslen, M., Kochi, K., and Lehmann, D. (2002). Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. *Methods and findings in experimental and clinical pharmacology*, 24 Suppl C:91–5.

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(1):49–65.

Pattwell, S. S., Duhoux, S., Hartley, C. A., Johnson, D. C., Jing, D., Elliott, M. D., Ruberry, E. J., Powers, A., Mehta, N., Yang, R. R., Soliman, F., Glatt, C. E., Casey, B. J., Ninan, I., and Lee, F. S. (2012). Altered fear learning across development in both mouse and human. *Proceedings of the National Academy of Sciences*.

Penny, W., Friston, K., Ashburner, J., Kiebel, S., and Nichols, T. (2007). *Statistical Parametric Mapping*. Elsevier.

Perego, P., Maggi, L., and Parini, S. (2009). BCI ++ : A new framework for Brain-Computer interface application. In *SEDE - ISCA*.

Phillips, C., Mattout, J., Rugg, M. D., Maquet, P., and Friston, K. J. (2005). An empirical Bayesian solution to the source reconstruction problem in EEG. *NeuroImage*, 24(4):997–1011.

Pi, H. J., Hangya, B., Kvitsiani, D., Sanders, J. I., Huang, Z. J., and Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature*.

Pieloth, C., Knosche, T. R., Maess, B., and Fuchs, M. (2014). Online Distributed Source Localization from EEG/MEG Data. *International Journal of Computing*, 13(1):17–24.

Pion-Tonachini, L., Makeig, S., and Kreutz-Delgado, K. (2017). Crowd labeling latent Dirichlet allocation. *Knowledge and Information Systems*, 53(3):749–765.

Pion-Tonachini, L., Sheng-Hsiou Hsu, Makeig, S., Tzyy-Ping Jung, Cauwenberghs, G., Hsu, S. H., Makeig, S., Jung, T. P., and Cauwenberghs, G. (2015). Real-time EEG Source-mapping Toolbox (REST): Online ICA and source localization. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, volume 2015-Novem, pages 4114–4117. IEEE.

Provost, F. and Fawcett, T. (2000). Robust Classification for Imprecise Environments. *Machine Learning*, 42(3):203–231.

Quitadamo, L. R., Abbafati, M., Saggio, G., Marciani, M. G., Cardarilli, G. C., and Bianchi, L. (2008). A UML model for the description of different brain-computer interface systems. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1363–1366. IEEE.

Radüntz, T., Scouten, J., Hochmuth, O., and Meffert, B. (2017). Automated EEG artifact elimination by applying machine learning algorithms to ICA-based features. *Journal of Neural Engineering*.

Reichel, L. and Sadok, H. (2008). A new L-curve for ill-posed problems. *Journal of Computational and Applied Mathematics*, 219(2):493–508.

Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., Bertrand, O., and Lécuyer, A. (2010). OpenViBE: An Open-Source Software Platform to Design, Test, and Use BrainâĂŞComputer Interfaces in Real and Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 19(1):35–53.

Robbins, H. (1956). An Empirical Bayes Approach to Statistics. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, (x):157–163.

Roger, C., Bénar, C. G., Vidal, F., Hasbroucq, T., and Burle, B. (2010). Rostral Cingulate Zone and correct response monitoring: ICA and source localization evidences for the unicity of correct- and error-negativities. *NeuroImage*, 51(1):391–403.

Sagi, Y., Tavor, I., Hofstetter, S., Tzur-Moryosef, S., Blumenfeld-Katzir, T., and Assaf, Y. (2012). Learning in the Fast Lane: New Insights into Neuroplasticity. *Neuron*.

Saha, S., Rana, R., Nesterets, Y., Tahtali, M., de Hoog, F., and Gureyev, T. (2017). Evaluating the performance of BSBL methodology for EEG source localization on a realistic head model. *International Journal of Imaging Systems and Technology*, 27(1):46–56.

Sanders, J. I. and Kepecs, A. (2014). A low-cost programmable pulse generator for physiology and behavior. *Frontiers in Neuroengineering*.

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., and Wolpaw, J. R. (2004). BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*.

Schlögl, A. and Brunner, C. (2008). BioSig: A free and open source software library for BCI research. *Computer*.

Sharp, P. E., Blair, H. T., and Cho, J. (2001). The anatomical and computational basis of the rat head-direction cell signal. *Trends in Neurosciences*, 24(5):289–294.

Siegle, J. H., López, A. C., Patel, Y. A., Abramov, K., Ohayon, S., and Voigts, J. (2017). Open Ephys: An open-source, plugin-based platform for multichannel electrophysiology. *Journal of Neural Engineering*.

Soliman, F., Glatt, C. E., Bath, K. G., Levita, L., Jones, R. M., Pattwel, S. S., Jing, D., Tottenham, N., Amso, D., Somerville, L. H., Voss, H. U., Glover, G., Ballon, D. J., Liston, C., Teslovich, T., Van Kempen, T., Lee, F. S., and Casey, B. J. (2010). A

genetic variant BDNF polymorphism alters extinction learning in both mouse and human. *Science*.

Stern, H. S. (2016). A Test by Any Other Name: P Values, Bayes Factors, and Statistical Inference. *Multivariate Behavioral Research*.

Susila, I. P., Kanoh, S., Miyamoto, K. I., and Yoshinobu, T. (2010). xBCI: A generic platform for development of an online BCI system. *IEEJ Transactions on Electrical and Electronic Engineering*.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, 2011.

Tamburro, G., Fiedler, P., Stone, D., Haueisen, J., and Comani, S. (2018). A new ICA-based fingerprint method for the automatic removal of physiological artifacts from EEG recordings. *PeerJ*, 6:e4380.

Tayeb, Z., Waniek, N., Fedjaev, J., Ghaboosi, N., Rychly, L., Widderich, C., Richter, C., Braun, J., Saveriano, M., Cheng, G., and Conradt, J. (2018). Gumpy: a Python toolbox suitable for hybrid brainâĂŞcomputer interfaces. *Journal of Neural Engineering*, 15(6):065003.

Terraneo, A., Leggio, L., Saladini, M., Ermani, M., Bonci, A., and Gallimberti, L. (2016). Transcranial magnetic stimulation of dorsolateral prefrontal cortex reduces cocaine use: A pilot study. *European Neuropsychopharmacology*, 26(1):37–44.

Thut, G. and Pascual-Leone, A. (2010). A review of combined TMS-EEG studies to characterize lasting effects of repetitive TMS and assess their usefulness in cognitive and clinical neuroscience.

Tikhonov, A. N. and Arsenin, V. Y. (1978). Solutions of Ill-Posed Problems. *Mathematics of Computation*, 32:1320–1322.

Tipping, M. E. (2001). Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*, 1:211–245.

Tomioka, R., Suzuki, T., and Sugiyama, M. (2009). Super-Linear Convergence of Dual Augmented-Lagrangian Algorithm for Sparsity Regularized Estimation. *Journal of Machine Learning Research*, 12:51.

Treder, M. S., Bahramisharif, A., Schmidt, N. M., Van Gerven, M. A., and Blankertz, B. (2011). Brain-computer interfacing using modulations of alpha activity induced by covert shifts of attention. *Journal of NeuroEngineering and Rehabilitation*.

Trujillo-Barreto, N. J., Aubert-Vázquez, E., and Penny, W. D. (2008). Bayesian M/EEG source reconstruction with spatio-temporal priors. *NeuroImage*, 39(1):318–335.

Trujillo-Barreto, N. J., Aubert-Vázquez, E., and Valdés-Sosa, P. A. (2004). Bayesian model averaging in EEG/MEG imaging. *NeuroImage*, 21(4):1300–1319.

UCSD Neural Engineering and Translation Labs (2018a). LSLOut Open Ephys plugin.

UCSD Neural Engineering and Translation Labs (2018b). SimBSI's Cortex Viewer.

UCSD Neural Engineering and Translation Labs (2018c). SimBSI's Scalp Viewer.

UCSD Swartz Center for Computational Neuroscience (2011). Lab Streaming Layer.

UCSD Swartz Center for Computational Neuroscience (2018a). LabRecorder.

UCSD Swartz Center for Computational Neuroscience (2018b). LSL Supported Devices.

Valdés-Hernández, P. A., von Ellenrieder, N., Ojeda-Gonzalez, A., Kochen, S., Alemán-Gómez, Y., Muravchik, C., and Valdés-Sosa, P. A. (2009). Approximate average head models for EEG source imaging. *Journal of Neuroscience Methods*, 185(1):125–132.

Valdés-Sosa, P. A., Vega-Hernández, M., Sánchez-Bornot, J. M., Martínez-Montes, E., and Bobes, M. A. (2009). EEG source imaging with spatio-temporal tomographic nonnegative independent component analysis. *Human Brain Mapping*, 30(6):1898–1910.

Van De Ville, D., Britz, J., and Michel, C. M. (2010). EEG microstate sequences in healthy humans at rest reveal scale-free dynamics. *Proceedings of the National Academy of Sciences*.

Van Der Maaten, L. and Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*.

Van Veen, B. D., van Drongelen, W., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*.

Vega-Hernández, M., Martínez-Montes, E., Sanchez-Bornot, J. M., Lage-Castellanos, A., and Valdés-Sosa, P. A. (2008). Penalized least squares methods for solving the EEG inverse problem. *Statistica Sinica*, 18(4):1535–1551.

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.

Vocat, R., Pourtois, G., and Vuilleumier, P. (2008). Unavoidable errors: A spatio-temporal analysis of time-course and neural sources of evoked potentials associated with error processing in a speeded task. *Neuropsychologia*, 46(10):2545–2555.

Wagner, J., Makeig, S., Gola, M., Neuper, C., and Muller-Putz, G. (2016). Distinct Band Oscillatory Networks Subserving Motor and Cognitive Control during Gait Adaptation. *Journal of Neuroscience*, 36(7):2212–2226.

Wibral, M., Vicente, R., and Lizier, J. T., editors (2014). *Directed Information Measures in Neuroscience*. Understanding Complex Systems. Springer Berlin Heidelberg, Berlin, Heidelberg.

Widmann, A. and Schröger, E. (2012). Filter Effects and Filter Artifacts in the Analysis of Electrophysiological Data. *Frontiers in Psychology*.

Widmann, A., Schröger, E., and Maess, B. (2015). Digital filter design for electrophysiological data - a practical approach. *Journal of Neuroscience Methods*.

Wipf, D. and Nagarajan, S. (2009). A unified Bayesian framework for MEG/EEG source imaging. *NeuroImage*, 44(3):947–966.

Wipf, D. P. and Nagarajan, S. S. (2008). A New View of Automatic Relevance Determination. In *Advances in Neural Information Processing Systems*, volume 20, pages 1625–1632.

Wipf, D. P., Owen, J. P., Attias, H. T., Sekihara, K., and Nagarajan, S. S. (2010). Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG. *NeuroImage*, 49(1):641–655.

Yamashita, O., Galka, A., Ozaki, T., Biscay, R., and Valdes-sosa, P. (2004). Recursive Penalized Least Squares Solution for Dynamical Inverse Problems of EEG Generation. *Human Brain Mapping*, 235(4):221–235.

Zhang, Z. and Rao, B. D. (2011). Sparse Signal Recovery With Temporally Correlated Source Vectors Using Sparse Bayesian Learning. *IEEE Journal of Selected Topics in Signal Processing*, 5(5):912–926.

Zhang, Z. and Rao, B. D. (2013). Extension of SBL Algorithms for the Recovery of Block Sparse Signals With Intra-Block Correlation. *IEEE Transactions on Signal Processing*, 61(8):2009–2015.

Ziegler, D. A., Mishra, J., and Gazzaley, A. (2015). The Acute and Chronic Impact of Technology on our Brain. In *The Wiley Handbook of Psychology, Technology, and Society*, pages 1–19. John Wiley & Sons, Ltd, Chichester, UK.

Zrenner, C., Belardinelli, P., Müller-Dahlhaus, F., and Ziemann, U. (2016). Closed-Loop Neuroscience and Non-Invasive Brain Stimulation: A Tale of Two Loops. *Frontiers in Cellular Neuroscience*.

Zrenner, C., Desideri, D., Belardinelli, P., and Ziemann, U. (2018). Real-time EEG-defined excitability states determine efficacy of TMS-induced plasticity in human motor cortex. *Brain Stimulation*.

Zrenner, C., Eytan, D., Wallach, A., Thier, P., and Marom, S. (2010). A Generic Framework for Real-Time Multi-Channel Neuronal Signal Analysis, Telemetry Control, and Sub-Millisecond Latency Feedback Generation. *Frontiers in Neuroscience*, 4.