Algorithms for Enhanced Spatiotemporal Imaging of Human Brain Function

by

Pavitra Krishnaswamy

B.S., University of Southern California (2007)
S.M., Massachusetts Institute of Technology (2010)

Submitted to the Harvard-MIT Division of Health Sciences and Technology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical and Medical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2014

Author .....
Harvard-MIT Division of Health Sciences and Technology
Aug 31, 2014

Certified by .....
Emery N. Brown, M.D., Ph.D.
Edward Hood Taplin Professor of Medical Engineering, MIT
Professor of Computational Neuroscience, MIT
Warren M. Zapol Professor of Anaesthesia, Harvard Medical School
Thesis Supervisor

Certified by .....
Patrick L. Purdon, Ph.D.
Assistant Professor of Anaesthesia, Harvard Medical School
Associate Bioengineer, Massachusetts General Hospital
Thesis Supervisor

Accepted by .....
Emery N. Brown, M.D., Ph.D.
Director, Harvard-MIT Program in Health Sciences and Technology
Professor of Computational Neuroscience and Health Sciences and Technology

Algorithms for Enhanced Spatiotemporal Imaging of Human Brain Function

by

Pavitra Krishnaswamy

Submitted to the Harvard-MIT Division of Health Sciences and Technology
on Aug 31, 2014, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical and Medical Engineering

Abstract

Studies of human brain function require technologies to non-invasively image neuronal dynamics with high spatiotemporal resolution. The electroencephalogram (EEG) and magnetoencephalogram (MEG) measure neuronal activity with high temporal resolution, and provide clinically accessible signatures of brain states. However, they have limited spatial resolution for regional dynamics. Combinations of M/EEG with functional and anatomical magnetic resonance imaging (MRI) can enable jointly high temporal and spatial resolution. In this thesis, we address two critical challenges limiting multimodal imaging studies of spatiotemporal brain dynamics.

First, simultaneous EEG-fMRI offers a promising means to relate rapidly evolving EEG signatures with slower regional dynamics measured on fMRI. However, the potential of this technique is undermined by MRI-related ballistocardiogram artifacts that corrupt the EEG. We identify a harmonic basis for these artifacts, develop a local likelihood estimation algorithm to remove them, and demonstrate enhanced recovery of oscillatory and evoked EEG dynamics in the MRI scanner.

Second, M/EEG source imaging offers a means to characterize rapidly evolving regional dynamics within an estimation framework informed by anatomical MRI. However, existing approaches are limited to cortical structures. Crucial dynamics in subcortical structures, which generate weaker M/EEG signals, are largely unexplored. We identify robust distinctions in M/EEG field patterns arising from subcortical and cortical structures, and develop a hierarchical subspace pursuit algorithm to estimate neural currents in subcortical structures. We validate efficacy for recovering thalamic and brainstem contributions in simulated and experimental studies. These results establish the feasibility of using non-invasive M/EEG measurements to estimate millisecond-scale dynamics involving subcortical structures.

Finally, we illustrate the potential of these techniques for novel studies in cognitive and clinical neuroscience. Within an EEG-fMRI study of auditory stimulus processing under propofol anesthesia, we observed EEG signatures accompanying distinct changes in thalamocortical dynamics at loss of consciousness and subsequently, at deeper levels of anesthesia. These results suggest neurophysiologic correlates to better interpret clinical EEG signatures demarcating brain dynamics under anesthesia.

Overall, the algorithms developed in this thesis provide novel opportunities to non-invasively relate fast timescale measures of neuronal activity with their underlying regional brain dynamics, thus paving a way for enhanced spatiotemporal imaging of human brain function.

Thesis Supervisor: Emery N. Brown, M.D., Ph.D.
Title: Edward Hood Taplin Professor of Medical Engineering, MIT
Professor of Computational Neuroscience, MIT
Warren M. Zapol Professor of Anaesthesia, Harvard Medical School

Thesis Supervisor: Patrick L. Purdon, Ph.D.
Title: Assistant Professor of Anaesthesia, Harvard Medical School
Associate Bioengineer, Massachusetts General Hospital

Acknowledgements

I feel grateful for the incredible privilege to work with and learn from many role models, teachers and colleagues through my graduate studies.

First and foremost, I would like to express my sincere gratitude to my thesis advisors, Emery Brown and Patrick Purdon. Emery's mentorship has been a cornerstone of my PhD experience, helping me to grow and develop as I have needed to throughout the past six years. Early on in my graduate career, he placed faith in my abilities, initiated collaborations and supported me as I independently defined and pursued a project that was beyond the lab's immediate focus. Later, when I needed to learn from and develop core expertise, he granted me the unique opportunities and vast research resources that have led to this thesis. His statistical expertise, biophysical intuition, and scientific insight have been a continued source of inspiration. I cannot thank him enough for all the guidance, support and freedom to explore and learn. Patrick's support, innovative ideas and enthusiasm have been really key to shaping my PhD research. I have benefited immensely from his guidance during our many brainstorming sessions, his continued optimism and focus during the hardest projects in this thesis, and his valuable perspectives on life and research. This thesis would not have turned out the way it did if not for intellectual inspiration from and access to the anesthesia datasets he has collected over the years. I feel very lucky to have had the chance to work with him, and will always be grateful for the opportunities he opened up for me.

I would also like to express my sincere gratitude to Matti Hamalainen for his mentorship, particularly on the source imaging project. This seemingly impossible project would have been truly impossible without his vision, teaching, collaboration, and the vast research resources that he enabled me to access. I have learned so much from his physically rooted approach to imaging and his ability to communicate complex ideas elegantly and effectively. I am also grateful for his continued feedback as part of my thesis committee, and for periodically sharing delightful Finnish chocolates, keeping my spirits up through the highs and lows of this thesis.

I am grateful to Elfar Adalsteinsson for his time and interest in serving on my thesis committee. His big picture perspectives and guidance have helped bring dimension to this work.

The collaborators I have had the opportunity to work with have played a vital role in realizing these projects. On the EEG-fMRI front, I am thankful to Giorgio Bonmassar for teaching me about EEG-fMRI experiments, sharing validation data and helping understand it, and for detailed feedback on the manuscripts. To Catherine Poulsen for sharing evoked response data from the other coast. On the source imaging front, I am very grateful to Gabriel Obregon-Henao for collaborating with me on the subspace pursuit implementation, teaching me to use software tools for data processing, spending many days and phone calls to refine source estimates, and patiently re-indexing our MATLAB code multiple times over. To Jyrki Ahveninen. Samantha Huang, Stephanie Rossi for enabling us to collect the M/EEG auditory evoked data without which we would not have been able to validate our source imaging approach. To Seppo Ahlfors and Tommi Raij for sharing their wonderful datasets for validation of our source imaging approach. On the propofol anesthesia front, I am thankful to a large team of MGH clinical investigators, led by Profs. Emery Brown, Patrick

Purdon, Eric Pierce and Giorgio Bonmassar, for acquiring this phenomenal EEG-fMRI data. I am fortunate to have had the chance to work on it. Special thanks to Dr. Pierce for his patience and time teaching me about anesthesia in the clinical setting during my clinical rotations.

These projects would not have been possible without the tireless software support from Colin Davey at Electrical Geodesics, and from Martin Luessi, Eugenio Iglesias, Koen Van Leemput, and the MNE and FreeSurfer core teams at the MGH Martinos Center.

Special thanks to Prof. Bruce Rosen for making himself accessible at regular student meetings. These have been a great source of feedback on proposals, and have helped connect me to wider initiatives at the MGH Martinos Center, and the larger world of brain imaging and clinical radiology.

All my colleagues in the vast Brown and Purdon labs have helped me in some way. In particular, thanks to Demba Ba, Behtash Babadi, Wasim Malik, Camilo Lamus, and Kevin Wong for many useful and enriching discussions on algorithm development, statistics, regression and compressed sensing. These have helped shape my computational work. To Sheri Leone for facilitating work and dealing with all the issues that have come up on a daily basis. To Katie Hartnack and Kara Pavone for managing clinical data collection and sharing, and all the paperwork that goes with it. To Laura Lewis and Rose Faghih for being fellow-graduate students in crime. To Michael Prerau for sharing amazing visualization tips. To Elvira Pirondini for sharing source estimation tips and code. To ShiNung Ching for guidance on navigating PhD-life. To Francisco Flores and Seong-Eun Kim for being fun office mates and tolerating my long phone calls to the Charlestown labs. To Christa Van Dort, Riccardo Barbieri, Seun Akeju, Patrick Stokes, and Veronica Weiner for their friendship and camaraderie. All of you have made the lab a fun place where I have looked forward to working each day. Outside the lab, I am also grateful to: Bruce Deffenbaugh for helpful discussions on exploratory data analysis, Sheraz Khan for helpful discussions on subcortical source estimation, and Robert Ajemian and Vincent Cheung for insightful neuroscience conversations.

The vast reservoir of scientific, clinical and engineering expertise at the Health Sciences and Technology program across MIT and Harvard Medical School have uniquely shaped my educational experience and will undoubtedly carry forward through my career. All the faculty, staff and students in the HST community have made my experience here enriching. Special thanks to my academic advisor Prof. Martha Gray, for helping me navigate the many challenges of graduate school. Thanks to Julie Greenberg, for helping me access a vast array of resources as needed. Thanks to Dr. Valerie Pronio-Stelluto for sharing her infectious enthusiasm for clinical medicine.

On a personal note, I am grateful for my immediate and extended family. Most importantly, I am forever indebted to my parents Mr. N. Krishnaswamy and Mrs. Mala Krishnaswamy for their continued sacrifices, encouragement, love, support, and unwavering commitment to my education and progress. To my father, I am truly sorry that I could not make this thesis happen in time for you to celebrate it with me. To my mother, your courage, will-power, and generosity have shaped my spirit, and your constant support has been crucial to my success. And to my husband, Anjan Soumyanaryanan, you are a real blessing in my life, making every day more meaningful. Your constant love and understanding have been an immense source of support and energy through my graduate education.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Understanding human brain function in cognition, behavior and illness remains an enduring challenge. This is because the brain is highly complex, with billions of neurons, interacting dynamically to receive, process, retrieve, transmit, and store information. Neurons signal via millisecond electrical impulses, communicate with other neurons in both local (mm-range) and distant (cm-range) regions. These interactions evolve instantaneously in response to external stimuli, drugs, or feedback from other brain regions, during creation and expression of memory, emotion and perception, across states of arousal, and over time due to plasticity, learning, development and aging. Thus, many scientific challenges in understanding brain function come about from the need to parse these complex dynamical interactions across diverse spatial and temporal scales. Further, it is technologically challenging to probe and interpret measures of these neuronal dynamics as they unfold in normal, clinical and disease states. As such, non-invasive techniques to access and understand spatiotemporal dynamics underlying human brain function are of wide interest.

1.1 Technologies for Imaging Brain Function

Understanding neuronal dynamics requires measuring neuronal activity at relevant spatial and temporal scales. Current human functional brain imaging technologies can be categorized as electromagnetic, biochemical and hemodynamic measures of neuronal activity (Fig. 1-1).

Electromagnetic measures of neuronal activity are direct indicators of neuronal currents, neuronal spiking or postsynaptic potentials, offer uniquely high temporal resolution on the order of milliseconds, and can be recorded either invasively or non-invasively [4–6]. Invasive electrocorticographic measurements (ECoG) in patient populations with surgically implanted intracranial electrodes (for recording from epidural and subdural surfaces or cortical gray matter) provide jointly high spatial and temporal resolution of regional neuronal dynamics. However, these electrodes are typically placed in selected (focal) regions limited to cortex, and are hard to obtain due to their

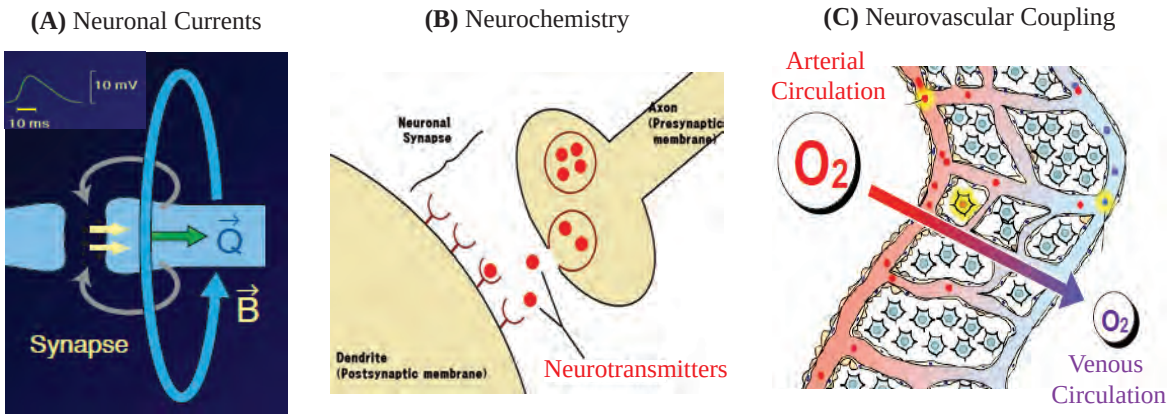

Figure 1-1: Biophysical Correlates of Neuronal Dynamics Accessible to Human Imaging. (*A*) Neuronal stimulation or signaling changes ionic permeability of the postsynaptic membrane, resulting in an ion current. This causes a measurable change in the postsynaptic membrane resting electric potential and also generates a neuromagnetic field. (*B*) Often, action potentials at an axon terminal trigger release of neurotransmitters, which function as chemical signaling messengers, and bind to postsynaptic membrane receptors, causing a measurable change in the receptor occupancy. (*C*) Neuronal activity extracts oxygen from cerebral blood supply, causing measurable changes in blood oxygenation. Figures adapted from [1–3].

invasive nature. Further, the clinical setting for these recordings limits the kinds of cognitive tasks and behaviors that can be accessed. Non-invasive electrophysiologic measurements obtained with electroencephalography (EEG) [4, 5] and magnetoencephalography (MEG) [7–9] offer more feasible alternatives and are thus widely used for human neuroscience studies. We will refer to MEG and/or EEG as M/EEG. However, these non-invasive techniques measure superpositions of fields generated by neuronal currents all over the brain, and thus cannot, by themselves, pinpoint regions where the currents originate [6]. Overall, electromagnetic measures offer uniquely high temporal resolution of neuronal currents, with easy clinical translation, but have limited spatial resolution and coverage (or spatial span) for regional dynamics.

Biochemical measures of neuronal activity provide insights into neuronal receptor function and cerebral metabolism, and can be obtained with nuclear medicine techniques such as positron emission tomography (PET) and single photon emission computed tomography (SPECT) [10, 11]. These techniques offer unique capabilities to study the neurochemistry associated with brain function, as well as glucose metabolism, with high sensitivity and controllable specificity based on radiotracer choice [12]. However, while these measures are linked (indirectly) to neuronal activity and can span the entire brain, reconstruction challenges limit the temporal and spatial resolution. Often, accurate reconstruction requires several minutes of recording, meaning low temporal resolution, and even so yields a low spatial resolution in the centimeter range [5, 13]. Further, the radiotracer requirement limits repeatability and application in healthy volunteers, and the need for local cyclotron facilities limits ease of clinical testing.

Hemodynamic measures of neuronal activity indicate regional changes in blood oxygenation fol-

lowing activation, and can be obtained with PET, optical techniques like near-infrared spectroscopy (NIRS) or magnetic resonance imaging techniques. As these methods are non-invasive and provide spatially resolved indicators of neuronal activity across the entire brain, they are widely used. The most significant hemodynamic brain imaging technique is Blood Oxygen Level Dependent functional MRI (BOLD-fMRI) [14–16]. This method is unique amongst hemodynamic measures as it has high spatial resolution, as well as sufficient temporal resolution to capture stimulus responses and network dynamics. Thus it has been the method of choice for a wide variety of functional neuroimaging studies over the last 20 years. However, as hemodynamic measures are only indirectly related to neuronal currents, they have two main drawbacks. First, the hemodynamic response reflects slow mass action. Specifically, the hemodynamic response to a brief millisecond-scale activation lags the actual activity by about 1-2 seconds and lasts for several seconds. Further, as the hemodynamic response reflects mass action, distinct neuronal dynamics such as excitation and inhibition, as well as feedforward vs. feedback processing may appear similar in PET or fMRI measurements [17]. Second, quantitative interpretation of hemodynamic signals mediated by neurovascular coupling remains a challenge. Specifically, it is difficult to relate the amplitude of fMRI signals to the magnitude of underlying neurophysiologic changes, independent of baseline physiology such as blood volume, oxygenation, flow or metabolism [18–20]. As such, fMRI findings are typically reported as statistical maps obtained via group averages across subjects. Thus, widespread clinical applications with diagnostic or prognostic utility in individual patients are yet to materialize.

Table 1.1 summarizes this review of human functional brain imaging modalities by grouping the biophysical correlates, resolutions for functional assessment, and spatial span.

Biophysical Correlates	**Imaging Modality**	**Spatial Resolution**	**Temporal Resolution**	**Spatial Span**
Neuronal Currents	ECoG	$0.05 - 5\,\text{mm}$	$1 - 5\,\text{msec}$	Focal Intracranial Structures
	MEG, EEG	$1 - 3\,\text{cm}$		Cortical Structures
Neurochemistry	PET, SPECT	$0.5 - 1\,\text{cm}$	$1\,\text{min}$	Full Brain
Neurovascular Coupling	PET	$0.5 - 1\,\text{cm}$	$1\,\text{min}$	Full Brain
	BOLD-fMRI	$1 - 3\,\text{mm}$	$5 - 6\,\text{sec}$	Full Brain
	NIRS	$3\,\text{cm}$	$5 - 8\,\text{sec}$	Superficial Cortical Structures

Table 1.1: Spatiotemporal Characteristics of Human Functional Neuroimaging Modalities. Resolution values collated from [5, 13, 17, 21–23]. Focal intracranial structures refers to (primarily cortical) regions where electrodes are implanted in a study. Cortical structures refers to neocortex. Spatial resolution for M/EEG are based on electromagnetic source imaging techniques (introduced in § 1.3)

As PET has low spatiotemporal resolution for functional assessments of brain dynamics, the rest of this introduction focuses on electromagnetic and hemodynamic measures of neuronal activity.

1.2 Imaging Studies and Spatiotemporal Dynamics

The availability of these functional brain imaging technologies has contributed significantly to our understanding of spatiotemporal dynamics underlying human brain function, reviewed in this section.

Electromagnetic measures of neuronal activity have enabled characterizations of (a) evoked responses in sensory, motor and cognitive processing [4], and (b) oscillatory brain dynamics in cognitive states like attention, memory and perception [13], and in clinically relevant states like sleep, anesthesia, and coma [24]. These measures have been used to characterize brain activity across frequency bands (theta, alpha, beta, gamma), and to associate activity in these bands with functional states (memory, idling, alertness, perception) [25–30]. High-density surface EEG and MEG studies have enabled quantitative characterizations of these states with metrics such as coherence and coupling across frequencies [4, 13]. Invasive ECoG cstudies in humans have enabled significant new understanding of these evoked and oscillatory dynamics in language, perception and sleep [31–33]. Together, these approaches have provided a variety of time-frequency electrophysiologic signatures that mark alterations in a variety of sensory, motor and higher cognitive functions.

However, despite these advances, the limited spatial resolution of these easily measured electrophysiologic signatures still remains a significant challenge. For example, questions concerning relatively simple dynamical responses such as evoked potentials under stimulus processing remain a matter of debate [34, 35]. Mechanisms and functional states underlying the largest and oldest known signatures on M/EEG, such as eyes closed alpha wave activity or sleep slow oscillations, are still to be understood [36–38]. Many open questions persist with regards to understanding the spectral dynamics and oscillatory structure governing large-scale neuronal interactions [39, 40]. Furthermore, while electrophysiology has immensely contributed to our understanding of cortical function, crucial dynamics involving subcortical structures are not typically addressed in humans. This is because electrodes are not typically implanted into subcortical structures in clinical patient populations, and surface M/EEG record small signals from these areas, making detection of activity in these areas difficult. In summary, due to the limited spatial dynamics accessible to human electrophysiology, much of what we know regarding the mechanisms and underlying brain dynamics in evoked and spontaneous states still comes from animal studies [41–43], and results can often be disconnected from human behavior.

This state of affairs has important clinical ramifications as well. First, questions relating to regional dynamics or sources of abnormal electrophysiologic signatures require invasive investigation. For example, in clinical conditions such as intractable epilepsy, the normal clinical course for localizing abnormal foci - despite all available non-invasive imaging techniques - still remains

surgical implantation of intracranial electrodes and invasive monitoring. Second, while electrophysiologic signatures hold potential diagnostic utility due to ease of clinical measurement and individual patient assessments, it remains unclear how to derive or interpret these signatures in cases they are not already known. For example, in states of altered arousal, awareness, perception, emotion or memory (e.g., coma, sleep, dreaming, pain, depression, or dementia), it would be desirable to have M/EEG biomarkers of dynamical changes within brain structures or circuits. However, as M/EEG signatures cannot be unambiguously related to regional dynamics (owing to low spatial resolution), obtaining clinically measurable biomarkers delineating specific regional or circuit-level changes remains challenging.

Hemodynamic measures of neuronal activity, such as fMRI recordings, have widely impacted scientific investigations across realms of neuroscience. For example, stimulus or task-induced response paradigms have enabled characterization of (a) sensory response and cognitive processing in behaviorally apparent and non-apparent brain states [44, 45], (b) working memory and memory encoding [46, 47], (b) anxiety, fear and other emotional states [48, 49], and (b) drug action, pain, and altered perception in clinically relevant states such as drug abuse, sleep and anesthesia, and psychiatric disease [50–53]. Further, functional connectivity assessments during non-task induced paradigms have enabled characterizations of regional and network dynamics in resting states, and have advanced knowledge of cognition, movement, memory, and unconsciousness [54–58].

However, despite these advances, the limited temporal dynamics available to fMRI, as well as the low SNR and the indirect nature of fMRI signals continue to pose challenges. For example, several questions prevail as to the physiology underlying fMRI-based connectivity measures and the degree to which they can be considered to reflect spatiotemporal neuronal dynamics beyond baseline metabolic correlations [17–20, 55]. Further, many challenges persist, across task-induced and resting-state studies, in quantitative interpretation of neurophysiologic changes underlying fMRI measurements and clinical translation of fMRI findings [17–20, 59]. Specifically, as most fMRI measures are statistical maps obtained across groups of subjects, the diagnostic utility of fMRI measures in individual patients is yet to take root in routine clinical practice [59].

In summary, electromagnetic measures (M/EEG) offer high temporal resolution and provide signatures of cognitive brain states that can be easily measured in the clinic. However, regional dynamics and brain mechanisms governing these signatures and in turn their underlying brain states are not well understood. On the contrary, hemodynamic measures (fMRI) resolve regional dynamics in many clinically or behaviorally non-apparent brain states. However, it would be desirable to obtain higher temporal resolution for these regional dynamics and determine clinically measurable markers for these states. Thus, relating fast, direct and clinically relevant measures of neuronal activity with their underlying regional dynamics is a long standing goal in neuroimaging. Such an endeavor would require simultaneously obtaining high temporal and spatial resolution. However, as the dynamics of brain function span spatial and temporal scales that are not accessible to any one imaging modality, synergistic combinations from different modalities are required.

1.3 Multimodal Techniques for Enhanced Resolution

Two sets of multimodal techniques are available for spatiotemporal brain imaging. First, simultaneous EEG and fMRI offers a means to relate fast EEG-based electrophysiologic signatures with concomitant spatially resolved but slower fMRI-based regional dynamics. This unique technique complements the temporal resolution of EEG with the spatial resolution of fMRI, and has immense potential to enable insights underlying the generation of EEG signatures as well as the regional dynamics relating to brain states characterized by the EEG changes [60–62]. Further, simultaneous measurements have significant advantages over combined analysis of separately acquired data – as simultaneous acquisition enables temporal correlations, ensures identical neurophysiologic events, environment and behavioral experience and minimizes confounds. As such, simultaneous EEG-fMRI has been applied to studies of the auditory steady state response [63], resting state (eyes closed) alpha rhythms [64–70], attention [71, 72], epilepsy [73, 74], and sleep and anesthesia [63, 75–78]. Yet, general utility of EEG-fMRI is limited by the immense challenges of recording and interpreting electrical signals within a hostile MRI environment (schematically illustrated in Fig. 1-2). Thus, the technique is yet to become a mainstay of functional neuroimaging.

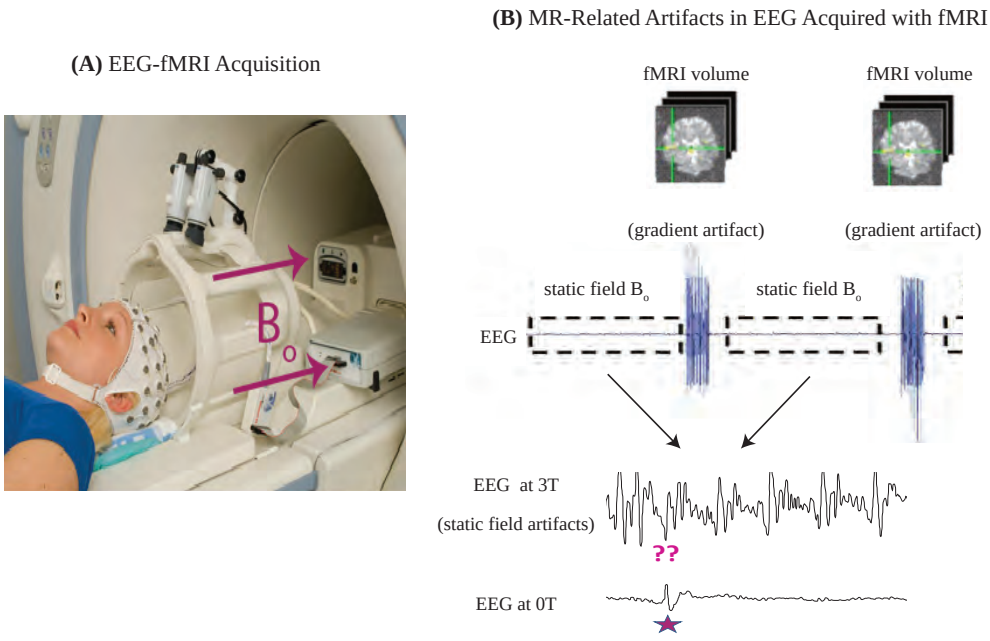

Figure 1-2: Illustration of Simultaneous EEG-fMRI. This technique enables the temporal association of millisecond-scale electrophysiologic signatures with slower but spatially resolved fMRI-based regional dynamics [60–62]. (A) High-density surface EEG can be acquired along with fMRI in a continuous or interleaved acquisition paradigm. B_o indicates magnitude of the static magnetic field. Figure adapted from [79]. (B) Illustration of interleaved acquisition paradigm, showing the two types of MR-related artifacts that corrupt EEG in the scanner. The large gradient artifact manifests during fMRI acquisition, and the static magnetic field ballistocardiogram artifact is continuously present. Bottom EEG time traces show challenging nature of the BCG artifact, which makes the known signature outside the scanner (0 T) marked by star difficult to detect inside the MRI scanner (3 T). Figures adapted from [63, 80].

Second, electromagnetic source imaging with non-invasive M/EEG recordings and MRI-based anatomic measures offers a means to probe regional dynamics at fast temporal scales [6, 21]. This technique involves computing solutions to the ill-posed electromagnetic inverse problem of recovering regional current distributions from M/EEG fields measured non-invasively at external sensors. The general framework is illustrated in Fig. 1-3.

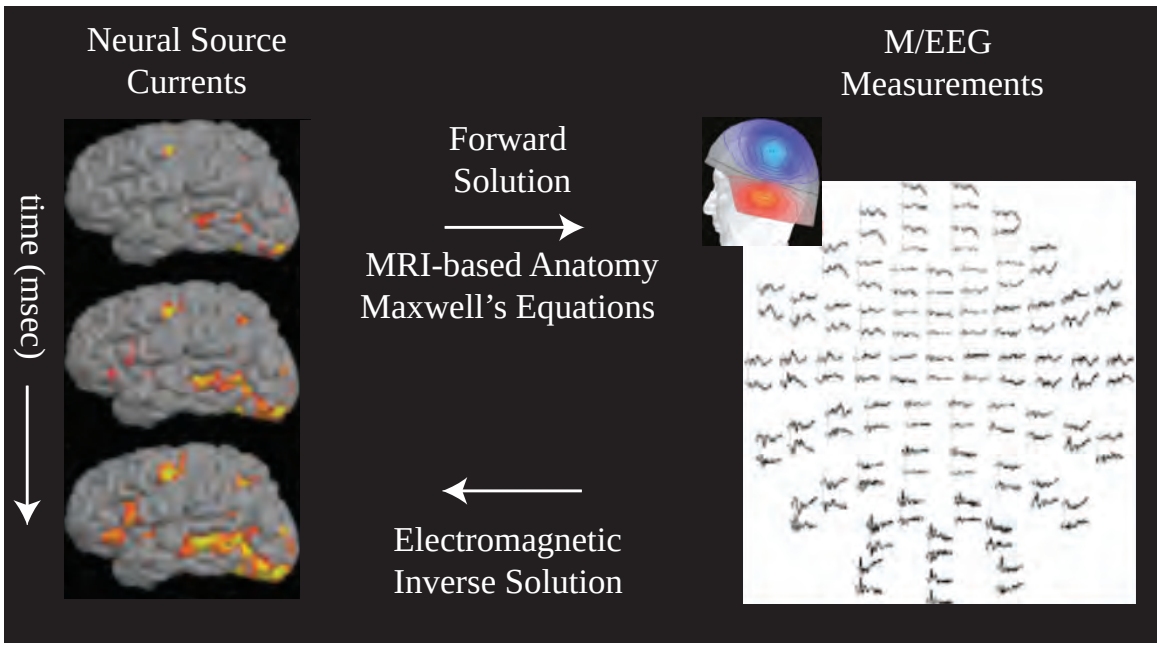

Figure 1-3: Illustration of Electromagnetic Source Imaging with M/EEG and Anatomical MRI. Regional neuronal currents (left) can be related to non-invasive M/EEG measurements (right) via the forward solution which employs anatomic knowledge from MRI and a numerical solution to Maxwell's equations. The measurements comprise field patterns across the M/EEG sensors (right, top corner), and evolve over time as the currents evolve over time (right, topography of measurement time courses). Given the M/EEG measurements and the forward solutions informed by anatomic MRI, it is possible to solve the electromagnetic inverse problem and obtain estimates of the regional source currents as they evolve. This technique enables the association of electrophysiologic signatures with their underlying regional dynamics at millisecond timescales. Figure adapted from [1, 81].

First, these approaches employ a current dipole as an elementary source, place these dipoles in brain structures based on anatomic measures from MRI, model neuronal activity within these brain structures as dipole currents, and relate these regional neuronal currents to non-invasive M/EEG measurements via a numerical solution of Maxwell's equations. This solution is termed the forward solution, and gives rise to the forward matrix. Then, by incorporating a model of noise, one arrives at a measurement equation. This measurement equation can then be inverted given the M/EEG data to obtain estimates of neuronal source currents across the brain. Thus, this technique is also often referred to as source estimation, source modeling or source localization [6, 21]. The development of accurate forward solutions that account for brain anatomy and cortical surface geometry [82, 83], along with significant advances in statistical signal processing, has led to reasonably reliable inverse solutions for cortical current distributions. These inverse estimates

can achieve resolutions as fine as $\sim 1\,\text{cm}$ [5, 13, 84–86]. As such, M/EEG source imaging has been applied widely to studies of cortical dynamics underlying evoked responses and oscillatory rhythms in sensory processing, attention, sleep, epilepsy, anesthesia and autism [37, 87–92]. However, due to biophysical factors governing the generation of MEG and EEG signals, the spatial span of sources that can be accessed via this technique is largely limited to cortical regions.

Overall, multimodal techniques offer unique opportunities for spatiotemporal brain imaging but are presently limited by a combination of biophysical and data interpretation challenges - reviewed in the following section.

1.4 Challenges in Simultaneous EEG-fMRI

Simultaneous EEG-fMRI offers a means to complement the unique information provided by EEG and fMRI. But recording and analyzing electrophysiologic data alongside fMRI poses a number of methodological challenges related to physics of the MRI environment [60, 62, 80, 93]. These challenges include (a) development of specialized hardware, (b) optimization of the acquisition paradigm, (c) effective artifact removal and data analysis, and (d) interpretation/integration. Over the past several years, significant progress on the first two challenges has been made [94] and a number of commercial systems for hardware and acquisition have been advanced – but the last two issues remain long-standing challenges [95, 96].

First, artifacts related to the MR-environment corrupt EEG recorded in the scanner and pose substantial challenges to the utility of this technique (Fig. 1-2). Two kinds of artifacts are encountered in EEG-fMRI studies [80]: (i) gradient-related artifacts arising during MRI acquisition and (ii) ballistocardiogram or cardiac pulsation artifacts arising from heart-beat and blood flow related head motion in the static field. While gradient-related artifacts can be relatively easily eliminated due to their predictable shape and timing [97], ballistocardiogram artifacts are highly challenging to remove as they vary over time and overlap true EEG signals substantially.

These artifacts pose difficulties for analysis of time-frequency signatures present in EEG obtained within the scanner. Common signatures like evoked responses, epileptic spikes, oscillatory dynamics (power spectra, phase measures, and cross-frequency coupling signatures [98, 99]), while easily observed outside the scanner, are hard to extract from artifactual EEG recorded within the scanner. While a number of techniques have attempted to address these artifacts [80, 100–102], they often do not allow reliable analysis of time frequency signatures of interest. These problems are exacerbated when the specific recording and scanning conditions are taken into account. For example, interleaved EEG-fMRI acquisition paradigms restrict EEG signals to short and discontinuous segments, wherein oscillatory dynamics are hard to analyze. Also, EEG-fMRI paradigms often involve long recording durations due to the need for sufficient spatial and temporal sampling [93], which often causes the EEG signal to degrade over time and makes assessments of task or time dependent variations difficult. As a consequence of the limited fidelity of EEG data acquired within

the scanner, EEG-fMRI techniques are yet to be adopted widely for studies of brain dynamics.

Second, statistical analyses correlating EEG and fMRI data need to be performed cautiously considering the biophysics and SNRs of the signals [103]. Analysis of fMRI changes must distinguish fMRI responses to stimulus or task paradigms, from responses correlated to electrophysiological signatures measured. Further, temporal correlations between EEG and fMRI must be tempered with an understanding of the frequency responses achievable in each of the two cases. Specifically, fMRI responses correlated to higher frequency EEG activity (such as gamma activity), must be evaluated carefully, as the neurovascular response does not allow fMRI changes at these frequencies. Moreover, in many cases activity on EEG can be correlated with a negative BOLD response, such as the responses seen during eyes closed alpha recordings, which can be hard to interpret [17]. In addition, while the fMRI responses can be used as priors for localizing sources of the EEG signals [81, 104, 105] – there are some complicating factors. For example, as the biophysics underlying EEG and fMRI measurements of neuronal activity are different, sources visible on EEG are not always visible in fMRI and vice versa [93]. For example, brief or highly synchronized neuronal currents may be seen in EEG but not in fMRI data, while activity within deep sources may present with a higher SNR in fMRI than in EEG. The differences in temporal resolutions and biophysics mentioned above, must also be carefully accounted for in performing these source analyses.

1.5 Challenges in Electromagnetic Source Imaging

Electromagnetic source imaging or source localization with M/EEG data seeks to estimate, with high temporal resolution, the source current distribution underlying externally measured fields based on (a) an electromagnetic generative model derived from Maxwell's equations and (b) anatomical information obtained via MRI. M/EEG generation physics suggests that measurements are primarily generated by post synaptic potentials in cortical pyramidal neurons, as subcortical sources are farther from the sensors and thus contribute relatively low amplitudes to non-invasive measurements [6, 106]. As a consequence, most source estimation methods only consider cortical source spaces. The applicability of this approach for deep brain sources remains questionable and the spatial span for studies of regional dynamics is currently limited to cortex.

However, it is well known that subcortical regions act synchronously with cortical regions to produce large oscillatory rhythms; work in tandem with cortical regions to process, suppress, and refine sensory input and motor actions; and dynamically interact with cortical regions in states of emotion, memory and arousal [107–111]. Accordingly, there has been a growing body of literature investigating whether deep sources can be localized from non-invasive M/EEG measurements [112–121]. These works use dipole fitting or classic minimum norm estimators [122] with discrete or distributed current dipole sources to characterize localizability of sources in hippocampus, basal ganglia, amygdala and thalamus. However, such efforts have found poor spatial resolution and accuracy for subcortical sources, requiring averaging over large numbers of trials to detect activity

in structures like the thalamus, especially in evoked settings [116, 117, 119]. As such, dynamics involving subcortical regions have been out of the purview of existing source imaging techniques.

To motivate improved inverse solutions, it is useful to discuss the reasons for difficulties in localizing subcortical sources. First, the depth of subcortical sources, along with the fact that electromagnetic fields drop substantially with distance to source currents, means that subcortical sources contribute weakly (low gain) to external sensors [6, 106]. While the low gain and SNR pose significant challenges, measurement amplitude is often not the only piece of information available. Many aspects of physiology are imaged with information contained in phase, distributions, or other amplitude-insensitive measures [123, 124]. However, these have not been explored in the subcortical electromagnetic inverse problem. Second, MEG is relatively insensitive to radial dipoles (in a perfect sphere radial dipoles produce no external magnetic field) and subcortical sources are often considered to be almost radial due to their deep location, close to the center of symmetry [6, 93]. However, the brain is not a perfect sphere. Further, we know that EEG does not lack sensitivity to radial source orientations. Therefore, combined M/EEG source imaging can address some of these issues. Third, many subcortical structures like thalamus are thought to have cytoarchitecture akin to a "closed-field", in that the electromagnetic fields they produce are confined within their volumes [93, 119]. However, no brain structure exactly obeys the geometric constraints of a closed field, and the spatial patterns of activity within these structures under physiologic conditions may correspond to configurations that allow the fields generated to project out of the structure. Further, even if the "closed field" argument holds for some structures or patterns of activity, sources in striatum, hippocampus and amygdala are thought to be open field and are believed to contribute significantly to non-invasive M/EEG data [119]. Thus, systematic characterizations considering (a) amplitude-insensitive information, (b) complementary information in MEG and EEG, and (c) physiologically relevant spatial configurations may motivate approaches to overcome challenges in subcortical source imaging.

1.6 Thesis Objectives

Overall, multimodal imaging approaches offer a framework for high resolution analyses of spatiotemporal dynamics underlying human brain function. They also afford the opportunity to relate non-invasive, fast, and direct electromagnetic measures of neuronal activity with underlying regional dynamics. However, two critical problems limit applicability of these approaches for studies of spatiotemporal dynamics in human neuroscience:

1. Relating non-invasive EEG signatures with spatially resolved regional dynamics measured on fMRI is highly challenging because MR-related ballistocardiogram artifacts corrupt the EEG.

2. Relating non-invasive M/EEG measurements with rapidly evolving regional or network dynamics in a source estimation framework is limited to cortical structures.

Thus, two priorities for effective spatiotemporal imaging with multimodal approaches are to (a) overcome MR-environment artifacts in EEG-fMRI and (b) extend the spatial span of M/EEG source estimation techniques to subcortical structures. These are the central objectives of this thesis. Achieving these goals promises a platform for resolving spatiotemporal brain dynamics with existing clinically translatable techniques. As present technologies cannot directly and non-invasively measure neuronal currents in the human brain, these objectives will open up novel lines of clinical and scientific investigation.

Achieving these thesis objectives requires overcoming fundamental biophysical and signal processing constraints with two unifying themes. First, both objectives aim to resolve weak signals of interest amidst substantially larger distractors which cannot be easily separated in time, frequency or spatial domains. Specifically, the EEG features of interest are much smaller than the MR-related artifacts and subcortical sources produce much smaller M/EEG signals than the cortical ones. While this reality paints a pessimistic picture of feasibility, the underlying biophysical features also offer optimism. Specifically, these problems do not involve completely unstructured information. Rather, they involve physiologically and physically generated signals which arise from structured dynamics and thus are bound to have natural structure. This structure, however, is unknown at the outset, and needs to be identified with systematic consideration of physiologic and physical features. Once the critical structure in the biophysical drivers and nature of the data is identified, however, it is possible to motivate and develop principled estimation and inference techniques. Second, it is desirable to achieve these objectives in computationally efficient ways that enable real-time and exploratory analyses on large volumes of data. Adapting algorithmic developments in statistics and signal processing for structured problems can vastly simplify implementations and enable efficient solutions. In summary, each problem in this thesis is addressed by understanding the physiology and physics underlying the data and the problem, then performing exploratory analyses on the data to identify key information and structure in the problem, translating this information into a mathematical framework or model, and finally developing an efficient and reliable solution tailored to the problem.

1.7 Thesis Overview

The remainder of the thesis is organized as follows. Chapter 2 reports on the identification of a harmonic basis for the EEG-fMRI ballistocardiogram artifact, and introduces a novel reference-free regression algorithm to reliably and efficiently overcome these artifacts. Chapter 3 reports on the identification of amplitude-insensitive distinctions in M/EEG field patterns arising from subcortical and cortical sources, and introduces a novel hierarchical subspace pursuit algorithm to estimate subcortical source currents from non-invasive M/EEG recordings. The potential of these advancements in spatiotemporal imaging for novel studies in cognitive and clinical neuroscience is illustrated in Chapter 4. Specifically, Chapter 4 reports on oscillatory EEG signatures accompanying simultaneous fMRI-based regional brain dynamics during induction and maintenance of

propofol general anesthesia. Taken together, Chapter 2-Chapter 4 offer a technological platform for enhanced multimodal imaging of spatiotemporal dynamics underlying human brain function, along with demonstrated applications for novel studies. Finally, Chapter 5 summarizes the thesis, and suggests future directions in the engineering, scientific and clinical realms.

Chapter 2

Removal of EEG-fMRI Ballistocardiogram Artifacts

Obtaining high quality electroencephalogram (EEG) data simultaneously with functional MRI (fMRI) recordings is increasingly relevant in cognitive and clinical neuroscience - as EEG-fMRI offers uniquely high spatiotemporal resolution for imaging brain activity. However, the utility of this technique is limited by ballistocardiogram (BCG) artifacts induced in the EEG by cardiac pulsation and head movement inside the magnetic field. In this chapter, we introduce a novel model-based harmonic regression algorithm to remove BCG artifacts from EEG recorded in the MRI scanner. We begin by identifying a harmonic basis for the BCG artifact. This allows us to model the artifact as a parametric function of this basis, and frame the problem within a classical time series parameter estimation context. We then develop an efficient maximum likelihood algorithm to identify model parameters, and estimate and subtract the BCG from corrupted EEG measurements. The modeling and estimation approach inherently accounts for temporal structure of interest in the brain generated EEG signals, and adapts to temporal variations in both artifact and signal. We test effectiveness on a common visual evoked potential paradigm, as well as on a challenging but practically relevant drug-induced oscillatory test case recorded at 3 T. In each case, we show that the harmonic regression algorithm effectively removes BCG artifacts, restores EEG signatures of interest within the scanner and outperforms existing techniques which rely on cardiac or motion reference signals. The algorithm enjoys three concurrent advantages – it does not require reference signals, performs in real-time, and works well even in low SNR situations. Thus, it is feasible for use in a range of EEG-fMRI paradigms and acquisition conditions. Overall, this chapter contributes a practical tool of relevance to a number of clinical, drug and cognitive neuroscience studies.

A preliminary version of this work was published in Krishnaswamy, Bonmassar, Purdon, and Brown, *Proceedings of the IEEE Engineering in Medicine and Biology Society* 2013: 5426-9 [125]. EEG recordings in the MRI scanner, used for validation of the algorithm, were obtained by Giorgio Bonmassar, Patrick Purdon, and Eric Pierce at the Massachusetts General Hospital, and Catherine Poulsen at Electrical Geodesics, Inc.

2.1 Introduction

Simultaneous recording of electroencephalogram (EEG) and functional MRI (fMRI) [60, 61] is a functional imaging technique of increasing relevance as it uniquely combines advantages and compensates disadvantages of standalone EEG and fMRI. On the one hand, standalone EEG measures brain activity directly, with high temporal resolution, and yields signature patterns (oscillations, evoked potentials, spikes) characterizing clinical and behavioral states - but it has low spatial resolution. On the other hand, standalone fMRI measures local neuronal activity across cortical and subcortical regions, with high spatial resolution, and resolves large-scale brain network changes - but it has low temporal resolution as it is an indirect (hemodynamic) indicator of neuronal activity. Simultaneous EEG-fMRI enhances fMRI by linking it with real-time EEG signatures [126], and enhances EEG by linking electrophysiologic features with fMRI-based regional dynamics [62]. As such, EEG-fMRI has applications for a wide variety of cognitive or clinical neuroscience studies including those pertaining to oscillatory rhythms in sleep-wake states, anesthesia, attention [64, 71, 75, 127, 128], evoked responses [129–133], and epileptic spikes [73, 74, 134]. However, realization of these applications is fundamentally limited by MR-environment related artifacts corrupting EEG recorded in the scanner.

The most challenging of the MR-environment artifacts is the ballistocardiogram (BCG) artifact, which is induced in the EEG by cardiac pulsation-related head motion in the static magnetic field [80, 135–137]. In typical recordings, large BCG artifacts (150-200 uV) overlap and obscure the true underlying EEG activity (5-100 uV) in both time and frequency domains- particularly for the 0-20 Hz band [80]. As a result, EEG recorded in the MRI scanner lacks (a) sensitivity for reliably detecting low amplitude EEG features like ERPs or spikes over ongoing artifact, and (b) specificity as it is easy to mistake BCG background for EEG theta or alpha rhythms or ictal discharge [100]. These problems compromise utility of EEG acquired in the scanner, and limit the robustness of quantitative EEG analyses like analysis of oscillatory coupling, scoring of sleep stages, detection of specific evoked potential components, or characterization of seizure discharge. Moreover, BCG artifacts scale with static field strength, posing increasing challenges given the emerging preference for high-field MRI techniques [80, 136, 137]. Thus, effective BCG artifact removal is crucial for EEG-fMRI applications, but is often difficult because of three concurrent challenges - (a) low SNR, (b) time-frequency overlap with signal, and (c) spatial inhomegeneities and unpredictable time variations in the artifact due to drifts in heart rate, blood pressure and pulsatile head motion.

Three types of techniques for removing BCG artifacts have been proposed: reference-based methods, basis decomposition techniques, and hybrid approaches integrating the two.

Reference-based methods use electrocardiogram (ECG) or motion sensor signals to specify templates for the BCG, and employ estimation methods using these templates to subtract out BCG artifacts. ECG-based subtraction techniques assume the BCG occurs only in a fixed part of the cardiac cycle, use ECG peaks to determine intervals when the BCG occurs, and average contam-

inated EEG segments across few such intervals to estimate the BCG waveform [61, 80, 138, 139]. While these methods are practical and widely used, they do not account for normal variations in timing, shape, and amplitude of the BCG artifact, resulting in out-of-phase BCG subtraction, systematic errors and large residuals [100, 102, 140]. Motion-based subtraction techniques assume a linear relation between head motion and BCG artifact, and employ adaptive filtering techniques to estimate the BCG waveform [100]. While this technique is rooted in physical features driving the BCG artifact, the adaptive filter formulation is (a) insensitive to nonlinearities in the motion-BCG relation, and (b) unsuitable for analyzing temporally structured features in the true brain EEG, which is treated as a white process. Moreover, these methods require robust reference signals suited for precise peak detection or adaptive filtering - but high magnetic fields often corrupt reference signals just as they corrupt the EEG signal.

To resolve these issues, reference-free basis decomposition approaches using ICA or wavelet transforms have been proposed [101, 141–143]. These techniques estimate independent time series or wavelet components constituting the BCG-corrupted EEG measurement, designate some of these components as artifactual, and selectively remove such components to subtract out the BCG. While the ICA assertion of statistical independence is justifiable given distinct origins of the brain-generated EEG signal and the pulse-driven BCG artifact, the substantial amplitude, time and frequency overlap between the two often renders unclear which basis components comprise "BCG artifact" and which components comprise "EEG signal". Thus, to designate artifact vs. signal, these methods employ significant post-processing and subjective case-specific criteria - which can bias artifact removal and render these techniques ineffective in fully separating BCG and EEG [95, 142, 144].

Finally, hybrid approaches integrate reference signals with basis decomposition methods [102, 145]. These studies use an ECG reference to identify artifact-corrupted intervals in the EEG [80], but estimate the BCG template using temporal principal component analysis or wavelet techniques in lieu of the moving average. Although this helps cope with some variations in the BCG waveform, and mitigates basis separability problems, difficulties in robustly identifying ECG peaks and BCG intervals persist.

Thus despite several advancements, prevailing challenges necessitate approaches that can address variations in the BCG waveform and deliver high SNR improvement without requiring high-quality reference signals or subjective separation criteria [95, 96].

In this chapter, we present a new reference-free model-based BCG removal algorithm to overcome these challenges. First, we specify a mathematical basis for the BCG that reflects the structured physical and physiologic features driving these artifacts. We then model the BCG time series as a parametric function of this basis. Thus defining the problem within a classical time series framework, we develop efficient likelihood-based regression techniques to identify model parameters that best explain the data, and then estimate and remove the BCG artifacts. Next, we

demonstrate the effectiveness of our technique on a common evoked potential test case as well as a challenging but practically relevant drug-induced oscillation test case acquired in the scanner. Finally, we benchmark performance under varied SNR and acquisition conditions in relation to existing methods, and discuss reasons for improved performance of our model-based approach.

2.2 Approach and Model

2.2.1 Physics and Physiology of BCG Artifacts

To identify a mathematical basis for the BCG, we first consider the physical and physiologic phenomena underlying these artifacts. BCG artifacts are believed to be generated primarily by heartbeat and blood flow related pulsatile motion of the head and electrodes in a static magnetic field [102, 146, 147]. This process is schematically illustrated in Fig. 2-1.

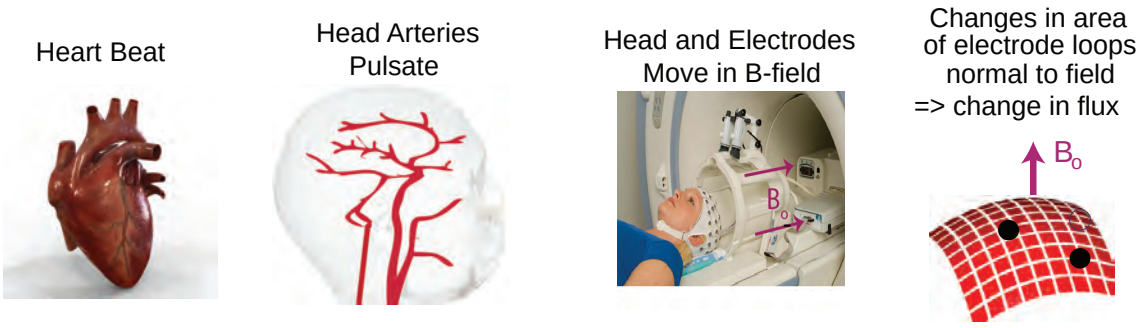

Figure 2-1: Physical and Physiologic Process Underlying BCG Artifacts. When the heart beats, the head arteries pulsate, causing movement of head and electrodes in the static magnetic field B_o. This, in turn, leads to changes in area of electrode loops perpendicular to B_o, causing changes in flux $\Phi = \int B_o \cdot dS$ and inducing (by Faraday's law) the ballistocardiogram (BCG) or 'pulse' artifacts. The induced artifacts add on to the true brain-generated EEG measurements. Variations in heart rate and blood flow over time, and non-uniformity of movement across the head surface cause time and spatial variations in the BCG respectively. Figures are collated and adapted from [79, 148, 149].

Thus the artifact is periodic in heart rate and driven by known, highly structured physical phenomena. Further, this view of the problem helps understand the large time and spatial variations in the artifact within the context of a structured process.

2.2.2 A Mathematical Basis for the BCG Artifact

To further understand the temporal structure in these artifacts, we inspect a variety of EEG recordings obtained in the MRI scanner. In Fig. 2-2*A*-*B*, we compare spectral structure of EEG obtained outside vs. inside the scanner under similar conditions. The spectrogram of EEG recorded outside the scanner clearly shows oscillatory EEG rhythms, however, inside the scanner these rhythms are obscured by comb-like harmonic streaks. As the underlying brain EEG state is similar in both cases, we note that the BCG is predominantly harmonic. Further, the BCG harmonics

manifest in time domain as comb-like pulsatile occurrences exhibiting the same periodicity as the heartbeat (Fig. 2-2*C*). Thus the BCG must comprise harmonics of the heart rate.

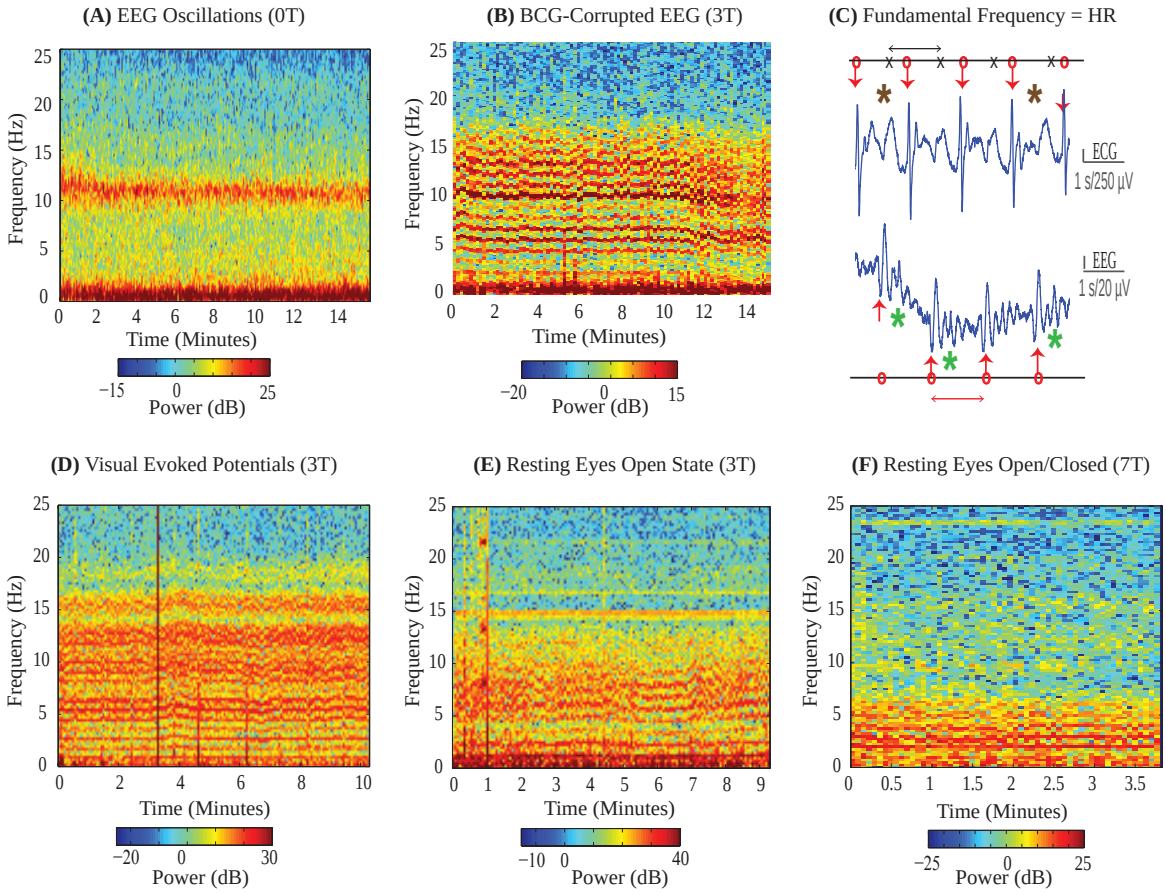

Figure 2-2: BCG is Harmonic in Heart Rate. EEG spectrograms recorded outside the scanner *(A)* and in 3 T MRI scanner *(B)* on 2 subjects undergoing similar levels of deep propofol anesthesia. *(A vs B)* BCG artifacts manifest as harmonic streaks in spectrogram, making it difficult to discern the true oscillatory structure seen outside the scanner. As spectrograms in *(A)* and *(B)* are obtained on different subjects, we can only contrast overall spectral structure and cannot infer absolute power comparisons. *(C)* Time series of BCG-corrupted EEG data (from Panel *B*) vs. simultaneous ECG recording. Successive high amplitude BCG combs (red arrows) have same periodicity as heart-beat. This same periodicity also manifests in the spectra as the frequency differential between adjacent harmonics in Panel *B*. Brown *'s indicate heart cycles where ECG R-wave peaks are hard to identify. Instances marked by green *'s have different relation between BCG and ECG peaks, showing the highly time-varying nature of the artifact. *(D-F)* Example spectra of BCG artifacts recorded in different experimental paradigms, on different subjects and different scanners. Gross motion artifacts appear as vertical lines in the spectrograms. Harmonic structure within the BCG artifact is independent of experimental conditions - therefore reflective of underlying physics and physiology. All spectra are computed using multitaper estimation with 0.3 Hz spectral resolution - high resolution spectral computations are necessary to discern structure in the BCG artifact as the harmonics are separated by heart rates in the ~ 1 Hz range. Data collection procedures are in [150] (Panel *A*), § 2.4.1 (Panels *B, C* and *D*), [93, 125] (Panel *E, F*).

Further, these observations generalize to different scanners, subjects and experimental conditions (Fig. 2-2*D-F*). Thus, they must represent fundamental physical and physiologic features of the BCG artifact, and suggest a natural harmonic basis for specifying the BCG artifact.

We note that the time domain traces in Fig. 2-2*C* show that the period, shape and intensity of the BCG combs varies over time. Also the BCG pulses comprise multiple high amplitude, discrete periodic pulse-like events, occurring across the cardiac cycle and modulating the true brain generated EEG signal - thus suggesting that the BCG is not limited to one time point in the cardiac cycle, but rather modulates the EEG signal throughout. Thus, the BCG is best specified as a parametric function of the harmonic basis, where the parameters capture variations within and across cardiac cycles.

Based on this analysis, we model the BCG artifact as a harmonic series with time-varying parameters, and also specify the temporal structure of interest in the brain-EEG signal parametrically. The parametric modeling approach defines artifact and signal templates without reference sensors, and allows objective separation between artifact and signal based on the character of the data without case-specific designations. We detail the model in the following section.

2.2.3 A Parametric Model

The EEG data comprises a multidimensional time series recorded using an array of sensors across the head, over K time points. The BCG-corrupted EEG measurement from a given sensor is denoted as $\mathbf{y} = [y_1, y_2, \ldots, y_K]'$. We model the measurement $\mathbf{y}$, from a given sensor, as a sum of the harmonic BCG artifact $\mathbf{s}$ and the true brain-generated EEG signal $\mathbf{v}$ at that sensor location. Then, for a given sensor, we have:

$$\mathbf{y} = \mathbf{s} + \mathbf{v}, \tag{2.1}$$

where $\mathbf{s} = [s_1, s_2, \ldots, s_K]'$ and $\mathbf{v} = [v_1, v_2, \ldots, v_K]'$.

The BCG artifact at a given time t is a harmonic series, which can be expressed as:

$$s_t = \mu_o + \mu_1 t + \sum_{r=1}^{R} A_r \cos(\omega r t) + B_r \sin(\omega r t), \tag{2.2}$$

where order R denotes the number of harmonics, $[\mu_o\ \mu_1]$ define a linear detrend term, $[A_r\ B_r]$ together define amplitude and phase of the r^{th} harmonic, and ω specifies the fundamental frequency. Based on Fig. 2-2*C*, ω is set by the heart rate.

Rewriting Eqn. 2.2 in matrix form, we have:

$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_K \end{bmatrix} = \begin{bmatrix} 1 & t_1 & \cos(\omega t_1) & \cdots & \sin(R\omega t_1) \\ 1 & t_2 & \cos(\omega t_2) & \cdots & \sin(R\omega t_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_K & \cos(\omega t_K) & \cdots & \sin(R\omega t_K) \end{bmatrix} \begin{bmatrix} \mu_o \\ \mu_1 \\ A_1 \\ \vdots \\ B_R \end{bmatrix}, \tag{2.3}$$

$$\mathbf{s} = \mathbf{Z}(\omega)\boldsymbol{\beta}, \tag{2.4}$$

where $\mathbf{Z}(\omega)$ and $\boldsymbol{\beta} = [\mu_o, \mu_1, A_1, B_1, A_2, \ldots, B_R]^{'}$ denote the harmonic basis and coefficients respectively.

The true brain EEG signal v_t can be modeled to reflect temporal features under study. A large variety of EEG signatures can be modeled with autoregressive (AR) processes. For example, spontaneous rhythms have inherent oscillatory or AR structure. Evoked responses, too, have a baseline EEG background that has AR structure. For epileptic spikes, the randomness associated with the baseline state can be modeled by a white (or zeroth order AR) process. In the general autoregressive case, at time t, we have:

$$v_t = \sum_{k=1}^{P} a_k v_{t-k} + \epsilon_t, \tag{2.5}$$

$$\epsilon_t \sim \mathcal{N}(0, \sigma^2), \tag{2.6}$$

where P denotes the order of the AR series $a_{k,t}$, and ϵ_t are independent identically distributed Gaussian random variables. Setting $P > 0$ corresponds to allowing temporal or spectral structure in the EEG, while $P = 0$ corresponds to modeling the EEG as a white process. This model form is specific enough to capture a variety of temporal or spectral features in the true brain EEG signal, yet general enough to not bias the algorithm towards predetermined periodicities or frequency bands.

Rewriting Eqn. 2.5 in Akaike Markovian matrix form, we have:

$$\begin{bmatrix} v_{t-1} \\ v_{t-2} \\ \vdots \\ v_{t-P} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_P \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} v_{t-2} \\ v_{t-3} \\ \vdots \\ v_{t-P-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{t-1} \\ \epsilon_{t-2} \\ \vdots \\ \epsilon_{t-P} \end{bmatrix}, \tag{2.7}$$

$$\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \tag{2.8}$$

where $\boldsymbol{\alpha} = [a_1, a_2, \ldots, a_P]^{'}$ and $\mathbf{Q}_{K\times K}$ denote the AR coefficients and the AR covariance matrix respectively. In the $P = 0$ white-EEG case, $\mathbf{Q}$ reduces to $\sigma^2\mathbf{I}_{K\times K}$, a multiple of the K-dimensional

identity matrix.

The composite model collating Eqn. 2.1, Eqn. 2.3, and Eqn. 2.7 can be written compactly as below:

$$\mathbf{y}\left(\omega,\, \boldsymbol{\beta},\, \boldsymbol{\alpha}, \sigma^{\mathbf{2}}\right) = \mathbf{Z}(\omega)\,\boldsymbol{\beta} + \mathbf{v}\left(\boldsymbol{\alpha}, \sigma^{\mathbf{2}}\right), \tag{2.9}$$

$$\mathbf{v}\left(\boldsymbol{\alpha}, \sigma^{2}\right) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}). \tag{2.10}$$

While the harmonic BCG and the autoregressive EEG models reflect the empirically observed overlap in BCG-EEG spectral and amplitude features, the overall forms of each model have distinct features that can appropriately decouple the BCG artifact from the true EEG signal. With this model, the problem of estimating and removing the BCG artifact becomes one of identifying parameters $\boldsymbol{u} = \left[\omega,\, \boldsymbol{\beta},\, \boldsymbol{\alpha},\, \sigma^2\right]$ that best explain the measurements $\mathbf{y}$.

2.3 Artifact Removal Algorithm

This parameter estimation problem is akin to a spectral estimation problem - except with a harmonic basis instead of frequencies spanning a continuum from 0 to $Fs/2$, adding on to a colored process instead of to white noise. However, as standard FFT-based spectral estimation procedures cannot differentiate the BCG from the AR process in this estimation problem, a time series estimation procedure is more relevant. Further, in order to resolve harmonics separated by $1 - 1.5\,\mathrm{Hz}$ (human heart rates) while retaining the underlying AR terms, high resolution estimates of the fundamental frequency ω are required. Also, as both the BCG artifact and the brain EEG signal vary over time (Fig. 2-2), the estimation procedure must account for time variations in model parameters defining artifact and signal. FFT-based techniques are limited in their ability to provide concomitantly high frequency and time resolution, and thus again, we turn to time series methods.

2.3.1 Maximum Likelihood Estimation

From Eqn. 2.9, it is evident that this is a problem of harmonic regression in the setting of colored noise. Maximum likelihood techniques are known to perform well in these classes of parametric estimation problems [151–154]. Further, these techniques can be easily adapted to the case of time varying parameters by exploiting the established local likelihood estimation framework [155]. Much like commonly used spectral estimation techniques, local likelihood estimation assumes quasi-stationarity of parameters within a short time segment, and applies maximum likelihood criteria to estimate model parameters on local time windows of data. As this is well suited to our problem, we develop a local maximum likelihood technique to estimate parameters defining the model in Eqn. 2.9. We describe the algorithm and implementation below for the general case of $P > 0$. A simplified version can be applied to the special case of $P = 0$.

For the i^{th} time window comprising T time points $t_i, t_{i+1}, \ldots t_{i+T-1}$, we denote the measurement

as a $T \times 1$ vector $\mathbf{y}_i$, and the local parameters as $\boldsymbol{u}_i = \left[\omega_i, \boldsymbol{\beta}_i, \boldsymbol{\alpha}_i, \sigma_i^2\right]$. Then, the local likelihood L_i of the unknown parameters given this measurement is:

$$L_i\left(\omega_i,\, \boldsymbol{\beta}_i,\, \boldsymbol{\alpha}_i,\, \sigma_i^2|\mathbf{y}_i\right) = L_i(\boldsymbol{u}_i|\mathbf{y}_i) \propto \left(\frac{1}{2\pi\sigma_i^2}\right)^{T/2} |\mathbf{Q}_i|^{-1/2} \exp\left(-\frac{S_{T,i}}{2\sigma_i^2}\right), \tag{2.11}$$

$$\text{where } S_{T,i} = \left(\mathbf{y}_i - \mathbf{Z}(\omega_i)\boldsymbol{\beta}_i\right)' \mathbf{Q}_i^{-1} \left(\mathbf{y}_i - \mathbf{Z}(\omega_i)\boldsymbol{\beta}_i\right). \tag{2.12}$$

Then, the optimal parameter estimates $\hat{\boldsymbol{u}}_i$ for the i^{th} window are those that maximize the local likelihood L_i:

$$\hat{\boldsymbol{u}}_i = \left[\hat{\omega}_i, \hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\alpha}}_i, \hat{\sigma}_i^2\right] = \arg\max_{\boldsymbol{u}_i} L_i(\boldsymbol{u}_i|\mathbf{y}_i). \tag{2.13}$$

Equivalently, the optimal parameter estimates $\hat{\boldsymbol{u}}_i$ for the i^{th} window minimize the negative log local likelihood:

$$-\log L_i\left(\omega_i, \boldsymbol{\beta}_i, \boldsymbol{\alpha}_i, \sigma_i^2|\mathbf{y}_i\right) = -\log L_i\left(\boldsymbol{u}_i|\mathbf{y}_i\right) \propto T\log(\sigma_i^2) + \log\left(|\mathbf{Q}_i|\right) + \frac{S_{T,i}}{\sigma_i^2}, \tag{2.14}$$

$$\hat{\boldsymbol{u}}_i = \left[\hat{\omega}_i, \hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\alpha}}_i, \hat{\sigma}_i^2\right] = \arg\min_{\boldsymbol{u}_i} \left(-\log L_i(\boldsymbol{u}_i|\mathbf{y}_i)\right). \tag{2.15}$$

The subscript i refers to the i^{th} time window, and all parameter estimates are denoted with hat's. For convenience, we will drop the subscript in $\mathbf{Q}$ and S_T.

2.3.2 Nested Cyclic Descent Implementation

This nonlinear and high-dimensional optimization can be solved by simplifying into a series of tractable problems that are amenable to efficient solutions. The minimization in Eqn. 2.14-Eqn. 2.15 can be written as a sequence of two minimizations:

$$\min_{\boldsymbol{u}_i}\left(-\log L_i\left(\boldsymbol{u}_i|\mathbf{y}_i\right)\right) = \min_{\omega_i}\left(\min_{\boldsymbol{\beta}_i,\boldsymbol{\alpha}_i,\sigma_i^2} -\log L_i\left(\boldsymbol{\beta}_i, \boldsymbol{\alpha}_i, \sigma_i^2|\omega_i, \mathbf{y}_i\right)\right) = \min_{\omega_i} C(\omega_i|\mathbf{y}_i), \tag{2.16}$$

where $C(\omega_i|\mathbf{y}_i)$ is the conditional likelihood given ω_i.

This conditional likelihood can be computed by solving the inner minimization and estimating $\hat{\boldsymbol{u}}(\omega_i) = \left[\hat{\boldsymbol{\beta}}(\omega_i), \hat{\boldsymbol{\alpha}}(\omega_i), \hat{\sigma}^2(\omega_i)\right]$:

$$C(\omega_i|\mathbf{y}_i) = \min_{\boldsymbol{\beta}_i,\boldsymbol{\alpha}_i,\sigma_i^2} -\log L_i\left(\boldsymbol{\beta}_i, \boldsymbol{\alpha}_i, \sigma_i^2\,|\omega_i, \mathbf{y}_i\right) \tag{2.17}$$

$$= -\log L_i\left(\omega_i, \hat{\boldsymbol{\beta}}(\omega_i), \hat{\boldsymbol{\alpha}}(\omega_i), \hat{\sigma}^2(\omega_i)\,|\mathbf{y}_i\right) \tag{2.18}$$

$$= T\log\left(\hat{\sigma}^2(\omega_i)\right) + \log\left(|\hat{\mathbf{Q}}(\omega_i)|\right) + \frac{\hat{S}_T(\omega_i)}{\hat{\sigma}^2(\omega_i)}, \tag{2.19}$$

$$\text{where } \hat{S}_T(\omega_i) = \left(\mathbf{y}_i - \mathbf{Z}(\omega_i)\hat{\boldsymbol{\beta}}(\omega_i)\right)' \hat{\mathbf{Q}}^{-1}(\omega_i) \left(\mathbf{y}_i - \mathbf{Z}(\omega_i)\hat{\boldsymbol{\beta}}(\omega_i)\right). \tag{2.20}$$

The terms $\hat{\boldsymbol{u}}(\omega_i)$, $\hat{\mathbf{Q}}(\omega_i)$, $\hat{S}_T(\omega_i)$ and therefore $C(\omega_i|\mathbf{y}_i)$ can be obtained by adapting a cyclic descent scheme [154, 156], which iterates between Generalized Least Squares, and a combination of Levinson-Durbin Recursions and the Burg algorithm. Then, the outer minimization in Eqn. 2.16 can be solved by minimizing $C(\omega_i|\mathbf{y}_i)$ across ω_i. The overall algorithm is illustrated in Fig. 2-3, and detailed below.

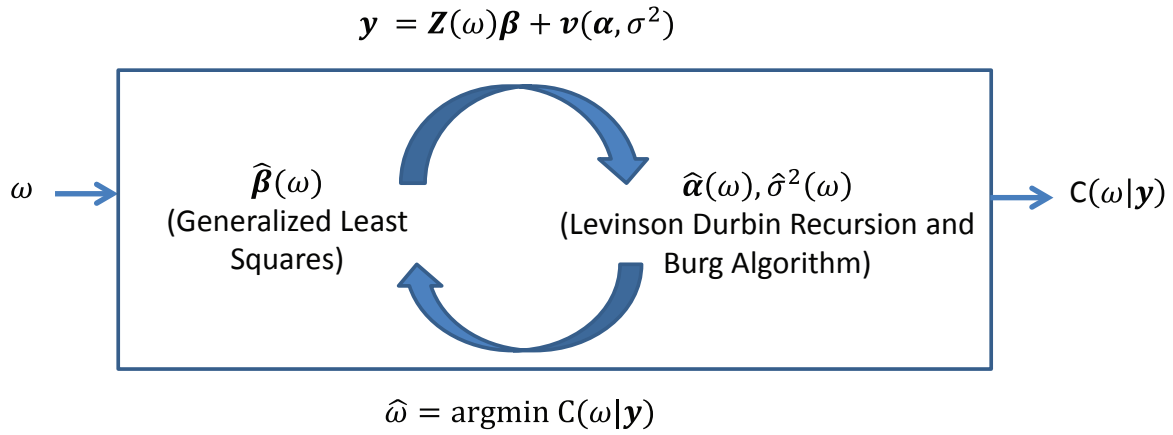

Figure 2-3: Harmonic Regression in Colored Noise to Estimate Model Parameters and Remove BCG Artifacts. For a given fundamental frequency ω, the conditional likelihood $C(\omega|\boldsymbol{y})$ can be computed using a cyclic descent scheme that estimates parameters $\left[\hat{\boldsymbol{\beta}}(\omega), \hat{\boldsymbol{\alpha}}(\omega), \hat{\sigma}^2(\omega)\right]$ by iterating between Generalized Least Squares for the harmonic amplitudes, and a combination of Levinson-Durbin Recursions and the Burg algorithm for the AR parameters. A one-dimensional optimization of $C(\omega|\boldsymbol{y})$ gives the optimal estimate of fundamental frequency $\hat{\omega}$. The corresponding $\left[\hat{\boldsymbol{\beta}}(\hat{\omega}), \hat{\boldsymbol{\alpha}}(\hat{\omega}), \hat{\sigma^2}(\hat{\omega})\right]$ are optimal estimates of the other parameters. This procedure can be performed in a local likelihood estimation framework by considering local segments in a moving window format. The overall algorithm extends work in [152–154, 156].

For a given ω_i:

1. Derive prior covariance matrix $\mathbf{W}$ for the $\boldsymbol{\beta}$ parameter (§ 2.8.1). The prior covariance helps account for background AR power in determination of harmonic amplitudes.

2. Initialize estimated inverse AR covariance $\hat{\mathbf{Q}}^{-1}(\omega_i)$ to identity $\mathbf{I}_{T\times T}$.

3. Obtain the maximum likelihood estimate of $\hat{\boldsymbol{\beta}}(\omega_i)$ via Generalized Least Squares on Eqn. 2.9:

$$\hat{\boldsymbol{\beta}}(\omega_i) = \left[\mathbf{Z}'(\omega_i)\hat{\mathbf{Q}}^{-1}(\omega_i)\,\mathbf{Z}(\omega_i) + \mathbf{W}^{-1}\right]^{-1} \mathbf{Z}'(\omega_i)\hat{\mathbf{Q}}^{-1}(\omega_i)\,\mathbf{y}_i.$$

4. Compute $\hat{S}_T(\omega_i)$ as defined in Eqn. 2.20.

5. Compute $\hat{\mathbf{v}}(\omega_i) = \mathbf{y}_i - \mathbf{Z}(\omega_i)\hat{\boldsymbol{\beta}}(\omega_i)$.

6. Estimate AR parameters $\hat{\boldsymbol{\alpha}}(\omega_i)$ and $\hat{\sigma}^2(\omega_i)$ underlying $\hat{\mathbf{v}}(\omega_i)$ using the Burg algorithm.

7. Update inverse covariance $\hat{\mathbf{Q}}^{-1}(\omega_i)$ of $\hat{\mathbf{v}}(\omega_i)$ using the Levinson-Durbin Recursion on AR model parameters estimated in step 6 [154].

8. Cycle through steps 3-7 until subsequent $\hat{\sigma}^2(\omega_i)$ values change $< 0.01\%$. Usually this converges within 4-6 iterations.

9. Compute $C(\omega_i|\mathbf{y}_i)$ as defined by Eqn. 2.17-Eqn. 2.20.

10. Return optimal parameter values $\hat{\boldsymbol{u}}(\omega_i)$, AR covariance estimate $\hat{\mathbf{Q}}(\omega_i)$, weighted mean square error estimate $\hat{S}_T(\omega_i)$ and likelihood cost $C(\omega_i|\mathbf{y}_i)$ for the given ω_i.

A one-dimensional optimization of $C(\omega_i|\mathbf{y}_i)$ over a range of plausible ω_i values gives the optimal estimate of the fundamental frequency:

$$\hat{\omega}_i = \arg\min_{\omega_i} C(\omega_i|\mathbf{y}_i). \tag{2.21}$$

Then, we have the optimal parameter estimates $\hat{\boldsymbol{u}}_i = \left[\hat{\omega}_i,\, \hat{\boldsymbol{\beta}}(\hat{\omega}_i),\, \hat{\boldsymbol{\alpha}}(\hat{\omega}_i),\, \hat{\sigma}^2(\hat{\omega}_i)\right]$ that solve Eqn. 2.14-Eqn. 2.15. In short, this is written as $\hat{\boldsymbol{u}}_i = \left[\hat{\omega}_i, \hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\alpha}}_i, \hat{\sigma}_i^2\right]$. It is noteworthy that these estimates carry optimality and convergence guarantees. A separation theorem [156] proves that the parameter estimates obtained via the cyclic descent procedure correspond to the global maximum of the likelihood function. In other words, if we find the optimal fundamental frequency $\hat{\omega}_i$, the associated Generalized Least Squares estimate $\hat{\boldsymbol{\beta}}_i$ and the associated AR parameter estimates $\hat{\boldsymbol{\alpha}}_i$ and $\hat{\sigma}_i^2$ are also optimal.

2.3.3 Algorithm Summary

Overall, the local likelihood estimation described here extends the harmonic regression literature - by providing a highly computationally feasible solution for the most general case of harmonic regression in colored noise. Most treatments of harmonic regression are limited to < 4 harmonics and a couple of AR terms, and require a known ω in the case of larger model orders [152–154]. Our algorithm enables a practical, efficient and optimal solution with a multiplicity of AR terms, a large number of harmonics, and unknown ω values. We proceed to test the method on experimental datasets.

2.4 Data Acquisition and Analysis

2.4.1 Data Acquisition

We test the algorithm on drug-induced oscillations and evoked response paradigms. Together our test cases represent a range of channel locations, SNR, EEG-fMRI recording paradigms, and acquisition systems.

2.4.1.1 Oscillatory Brain Activity

EEG oscillations were recorded in an interleaved EEG-fMRI study conducted at 3 T (Siemens, Erlangen, Germany) during induction and maintenance of propofol general anesthesia in $50-60$ y/o volunteers who were healthy post-tracheostomy [63]. As the EEG oscillations in this paradigm are concentrated in the fronto-temporal regions [150] where the BCG artifacts tend to be most severe, these recordings provide a challenging oscillatory test case for the algorithm. Continuous EEG was acquired at 950 Hz using an MR-compatible acquisition system configured per the International 10/20 layout, with electrodes placed in adjacent bipolar pairs at 24 locations across the head. EEG electrodes were made of Ag/Cl, housed in plastic (Gereonics Inc., Solana Beach, CA), bonded to carbon fiber wires ("Fiber-Ohm" wires with 7 ohms/inch resistance used to avoid RF heating during MRI; Marktek, Inc., Chesterfield, MI). Effective TR for fMRI acquisition was 9 seconds, and each fMRI volume was acquired in 1 second (EPI artifact periods of 1 second). The pulse was measured every 5 minutes. Piezoelectric motion sensor data [100] were simultaneously acquired with the same amplification and acquisition system as the EEG for comparisons between our reference-free approach and motion-based BCG removal techniques. This study was approved by the Partners Human Research Committee at Massachusetts General Hospital (protocol number 1999-P-010748), and all study subjects provided written informed consent.

2.4.1.2 Evoked Response Potentials

Evoked response EEG recordings were obtained on 2 healthy volunteers in a head-only 3 T scanner (Siemens Allegra) during a visual evoked response paradigm comprising 1150 trials of a vertical upper and lower checkerboard wedge stimulus. Continuous EEG was acquired at 250 Hz using an MR-compatible EEG acquisition system (Electrical Geodesics, Inc., Eugene, OR) with vertex-referenced electrodes placed at 256 locations across the head. The simultaneous EPI acquisition was turned off during the recording, thus the data contains no gradient artifacts beyond the static magnetic-field BCG artifact. The cryopump was also turned off during the recording. ECG data were simultaneously recorded with the same acquisition system as the EEG for comparisons between our reference-free approach and ECG-based BCG removal techniques. This study was approved by the Human Subjects Institutional Review Boards of Electrical Geodesics, Inc. and the University of Oregon, and all study subjects provided written informed consent.

2.4.2 Data Analysis

2.4.2.1 Pre-Processing

For simultaneous EEG-fMRI studies, the raw data is preprocessed with standard methods to remove gradient artifacts [97]. For interleaved studies, the data acquisition format naturally makes EEG windows corrupted by 'pure' BCG artifacts readily available for analysis. The length of the moving

window sets limits on the achievable frequency resolution. Further, as one set of parameters are estimated for each data window, the length of the moving window imposes a time resolution on how quickly the estimated parameters can change in response to drifts in the BCG artifacts and the true brain EEG. Thus, the moving window interval T is chosen to be long relative to both the fundamental harmonic and the slowest EEG frequency band of interest (so the estimation can use information across a few cycles), while being short enough to account for beat-to-beat drifts in heart rate and harmonic amplitudes. This tradeoff is nicely accounted with $2-6$ second data intervals for the test cases illustrated here (§ 2.8.2). Each channel of data is treated independently and in parallel for the data analyses. All analyses are performed on the native EEG referencing scheme used in the data collection. After preprocessing the data, we performed the following analysis steps for each test case under study.

2.4.2.2 Model Selection

Representative blocks across multiple channels and time periods in the data series are selected to assess suitable model orders. Then, harmonic orders R and AR orders P are jointly derived based on the uncorrected and corrected Akaike and Bayesian Information Criteria (AIC, AICc and BIC). For each selected window of data, and each possible pair of harmonic and AR orders (r, p), the residual variance σ^2 is estimated as detailed in § 2.2.3, and the information criteria are computed as below:

$$AIC = T\log\sigma^2 + 2(2r+p+1), \tag{2.22}$$

$$AICc = AIC + \frac{2(2r+p+1)(2r+p+2)}{T-(2r+p+2)}, \tag{2.23}$$

$$BIC = T\log\sigma^2 + (2r+p+1)\log(r). \tag{2.24}$$

Model order guidelines are selected by assessing the pairs of (r, p) that minimize the information criteria across the chosen representative blocks. For example, $R, P = \arg\min_{r,p} BIC$. This choice allows enough terms for goodness of fit while penalizing large numbers of parameters.

Standard goodness of fit assessments on the residual noise term $\hat{\boldsymbol{\epsilon}}$ are integrated into the model selection. First, we assess goodness of fit by comparing the normalized cumulative periodograms (NCP) of the residual noise series and an ideal white noise series of the same length. If the residual noise NCP is within 95 % whiteness bounds around the ideal white noise then we deem the temporal correlation in the residual to be insignificant. Further, we use quantile-quantile plot of the residuals to confirm Gaussianity. Model orders that lead to residuals not fulfilling these goodness of fit criteria are rejected regardless of the associated information criteria.

The model orders derived above are then checked against the additional constraints that (a) the harmonic order is high enough to cover all significant harmonic frequencies in the raw data power spectra, and (b) the AR order is of the same scale as the expected number of oscillatory features

(2 poles per spectral peak) in the brain EEG. Based on the sum total of these evaluations, one pair of harmonic and AR model orders (R, P) is selected to analyze the entire EEG recording under study. Typical choices of model orders are listed in § 2.8.2.

2.4.2.3 Parameter Estimation

The local likelihood estimation algorithm is implemented in MATLAB (Mathworks, Natick, MA). Inputs to the algorithm include the measured data series, the harmonic and AR model orders, and bounds on the possible fundamental frequency values for the final 1-D optimization in Eqn. 2.21. Denoting a subject's typical heart rate as H, the bounds are set as $[\min(40,\, 0.5H),\ \max(1.5H,\, 150)]$ beats per minute. One set of model parameter estimates $\hat{\boldsymbol{u}}_i$ is obtained for each data window i, while moving windows along to analyze the entire data series. As heart rate does not change drastically from one moving window to the next, the fundamental frequency estimate for the i^{th} window, ω_i, can be used to initialize the search for the fundamental frequency ω_{i+1} corresponding to the $(i+1)^{\text{th}}$ window. Similarly, as the EEG structure typically does not change drastically from one moving window to the next, the inverse AR covariance estimate for the i^{th} window can be used to initialize the AR covariance for the $(i+1)^{\text{th}}$ window (steps 2 and 10 of cyclic descent implementation in § 2.3.2). This has the effect of ensuring that parameter estimates evolve relatively smoothly from one window to the next. Once the parameters are estimated for a given data window, time series estimates of the BCG artifacts ($\hat{\mathbf{s}}_i$), clean true-brain EEG ($\hat{\mathbf{v}}_i$), and residual noise ($\hat{\boldsymbol{\epsilon}}_i$) are computed.

2.4.2.4 Computational Speed

The local likelihood estimation algorithm and the nested cyclic descent implementation are, by design, more efficient than a standard Newton's procedure with numerous gradient and Hessian computations. We perform three further simplifications to increase computational speed significantly. First, as the computation of $C(\omega_i|\mathbf{y}_i)$ is the most computationally demanding task, we downsample the data segments used during the search for $\hat{\omega}_i$ to enable faster computations, while retaining the fully sampled data segments for the final cyclic descent using the $\hat{\omega}_i$ to maintain high precision. Second, we speed up the Levinson-Durbin Recursions in [154] significantly by using a block matrix multiplication form to efficiently obtain $\hat{\mathbf{Q}}(\omega_{\mathbf{i}})$ from the Cholesky and diagonal factors. Third, we parallelize the cyclic descents performed during the search for $\hat{\omega}_i$. These practical features allow us to remove artifacts in a 5 second long data segment in < 4.5 seconds using an 8-core Intel Xeon$^{\text{TM}}$ workstation. Thus our method is well-suited for real time analysis during acquisition of large EEG-fMRI datasets – and is especially relevant for long duration studies or for stimulus/drug dosing paradigms that require dynamic information on the subject's brain state.

2.4.2.5 Comparisons with Existing Algorithms

We benchmark performance of our algorithm against three commonly used techniques - namely (a) motion-sensor based adaptive filtering [100], (b) ECG-based moving average subtraction [80] and (c) optimal basis separation (OBS) which uses ECG signals in combination with temporal PCA [102]. The motion-sensor-based adaptive filter is augmented with an E-M algorithm for model parameter estimation as in [93]. In case of interleaved recordings, the adaptive filtering is performed using intermittent observations within a missing data framework. The ECG-based subtraction and OBS algorithms are implemented within the FMRIB plug-in [102, 157] for EEGLAB [158], provided by the University of Oxford Centre for Functional MRI of the Brain (FMRIB).

2.5 Estimation Results

We validate our algorithm on the two experimental datasets from § 2.4.1. First, we present results on propofol-induced EEG rhythms recorded at 3 T. This is a challenging test case on two counts - (a) propofol oscillations involve frontal channels [150], that are known to represent very challenging BCG artifacts [80], and (b) propofol induces oscillations across multiple frequency bands [150]. We use a representative segment of oscillatory activity in a frontal channel (F4 →Fp2) to detail goodness of fit of the proposed model (Fig. 2-4), convergence characteristics of the proposed algorithm (Fig. 2-5) and demonstrate the ability to recover oscillatory EEG signatures of interest in the scanner (Fig. 2-6-Fig. 2-7). Second, we present results for a visual evoked response test case (Fig. 2-8) to demonstrate applicability to varied paradigms.

2.5.1 Model Selection and Goodness of Fit

Fig. 2-4*A*-*C* illustrate how the joint information criteria (§ 2.4.2.2) are used to determine harmonic and AR model orders (R, P) that adequately explain the structure in the data. As the signal is largely harmonic, we first determine optimal harmonic order R using the joint BIC (Fig. 2-4*B*), then fit the data to R harmonics, and determine AR orders P that best capture the remaining temporal structure (Fig. 2-4*C*). The best harmonic order is typically independent of the AR orders, as the BCG harmonics dominate the data. Usually, a range of suitable AR orders, denoted by the inflection points, can explain oscillations of interest. Adding more AR terms merely fits more background EEG power. These trends are insensitive to choice of information criteria in Eqn. 2.22-Eqn. 2.24.

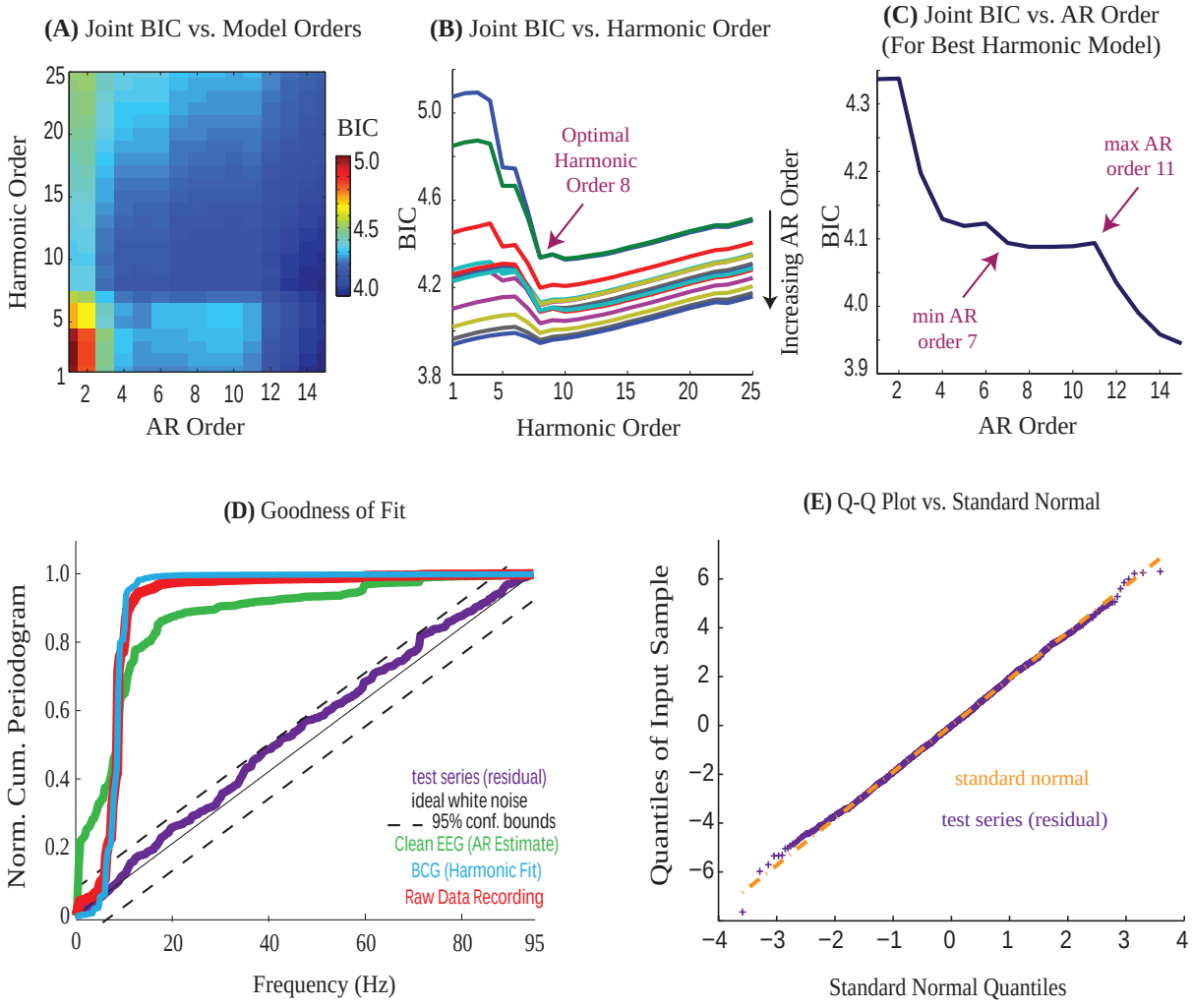

Figure 2-4: Model and Implementation Fulfill Appropriate Goodness of Fit Properties. Model order selection and goodness of fit are shown for a representative 3 second segment of propofol-induced frontal EEG rhythms recorded at 3 T. *(A)* Joint Bayesian Information Criterion (BIC) as a function of the harmonic and autoregressive model orders. *(B)* Joint BIC from *(A)* plotted as a function of harmonic orders - optimal harmonic order per the information criteria for this segment is 8. *(C)* Joint BIC for the optimal harmonic order from *(B)* plotted as a function of AR orders - the leveling off in the joint BIC for $7-11$ AR terms suggests that $7-11$ AR coefficients may be sufficient to explain this data. These criteria were assessed across several data segments in a recording to obtain one model order guideline (R_{bic}, P_{bic}) that most consistently explains the recorded data. In practice, the BIC results serve as a guide to minimum number of harmonics needed, and R is set to $2R_{\text{bic}}$ to ensure that higher harmonics are not ignored. These selected model orders were used to perform parameter estimation and assess goodness of fit. *(D-E)* Results of goodness of fit analysis on residuals $\boldsymbol{\epsilon}$ that remain after fitting with joint harmonic-AR model. *(D)* Normalized cumulative periodograms of the raw data (red), the harmonic fit (cyan), the AR estimate (green) and the residual series (purple). While the harmonic and AR components take on much of the low frequency power and temporal structure in the raw recording, the residual series $\boldsymbol{\epsilon}$ lies within 95 % whiteness bounds (dashed lines around the ideal white noise in black). *(E)* Q-Q plot of residual series $\boldsymbol{\epsilon}$ confirms Gaussianity, as expected by the model. Thus, the chosen model explains the data.

Fig. 2-4*D-E* illustrate goodness of fit results on residuals $\hat{\boldsymbol{\epsilon}}$ to show that the chosen model orders (R, P) indeed explain the data. While the harmonics explain much of the spectral structure in the raw EEG recording, the AR terms are necessary to capture all the systematic variance in the data (Fig. 2-4*D*). As the residuals $\hat{\boldsymbol{\epsilon}}$ under test lie within 95 % whiteness bounds (Fig. 2-4*D*) and have the modeled Gaussian structure (Fig. 2-4*E*), the proposed joint harmonic-AR model is a sufficiently good reflection of structure in the data, and thus can be used to estimate parameters.

2.5.2 Accuracy of Parameter Estimates

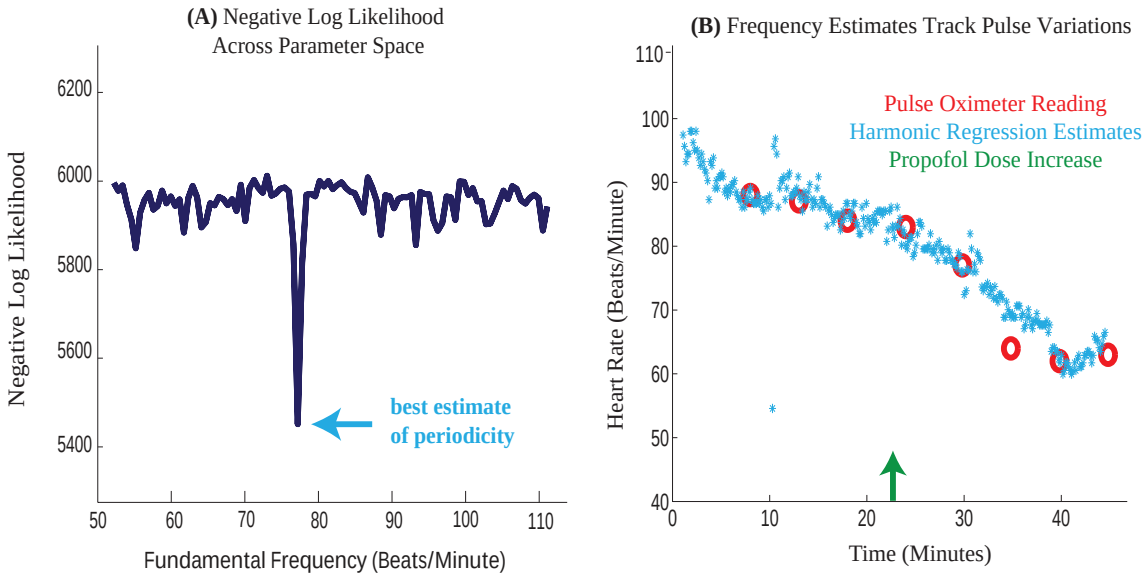

Figure 2-5: Parameter Estimates are Unique and Accurate. Data likelihood and parameter estimates are shown for the same segment of frontal EEG recordings during propofol anesthesia as is presented in Fig. 2-4. *(A)* The data likelihood, plotted across all possible fundamental frequencies, has a global minimum in fundamental frequency (77 beats per minute for this segment) - thus convergence of the algorithm is robust. *(B)* Estimated fundamental frequency ω at each time in the study. It is clear that algorithm's parametric estimates track the clinically recorded heart rates. The local likelihood implementation helps the algorithm adapt to changing parameters for the BCG artifact, even as drug dose increases (green arrow). As accurate frequency estimates lead to accurate estimates of other parameters [156], our algorithm provides accurate estimates of parameters underlying the BCG artifact and clean EEG data.

Fig. 2-5 illustrates the convergence characteristics of the proposed algorithm. Fig. 2-5*A* shows the conditional likelihood $C(\omega|\mathbf{y})$ for one segment of data, across the range of possible fundamental frequencies ω, binned at 0.01 Hz. It is evident that $C(\omega|\mathbf{y})$ exhibits a unique minimum in ω, suggesting that the final 1-D optimization converges robustly. We note that the shape of $C(\omega|\mathbf{y})$, and thus the choice of ω, are primarily influenced by the residual power (σ^2) and AR covariance ($\mathbf{Q}$). Values of ω that correctly account for the high amplitude harmonic template lead to lower

σ^2, $|\mathbf{Q}|$ and thus C. Further, the AR covariance $\mathbf{Q}$ helps account for the AR power and spectral structure, correcting the choice of ω so that the algorithm does not mistake brain-generated AR power for a harmonic artifact that needs to be removed. We note that the optimal ω is relatively robust to selection of model order. Typically, using models comprising a few harmonics and more than $4-5$ AR terms for the estimation leads to similar fundamental frequency values. As the harmonic structure is evident in the data, the uniqueness of this solution and its insensitivity to exact model order is not surprising.

Further, the nature of the conditional likelihood across ω enables two very practical features for our algorithm. First, even if the moving windows of data are short and do not contain many cycles of the first harmonic, they will contain many more cycles of the higher harmonics. Thus, the estimator can use the repeat cycles of higher harmonics to derive precisely the range of fundamental frequencies possible given the harmonic structure in the data. Therefore, even with moving windows as short as 3 seconds, we can obtain high resolution estimates of fundamental frequencies in the typical heart-rate range of $0.7-1.5$ Hz. This frequency resolution is significantly higher than can be obtained with a frequency-domain based spectral estimator on a short moving window, and is possible because the statistical estimation accounts for prior information in the harmonic structure. In other words, we can maintain high time resolution with short moving windows and quickly adapt to beat-to-beat changes in heart rates alongside high frequency resolution. Second, the consistent presence of global optima in the likelihood costs enables us to vastly simplify the 1-D optimization routine for real-time implementation. Thus, the structure in the data confers our algorithm with the triple advantages of high frequency resolution, high time resolution and real-time performance.

Fig. 2-5*B* shows that the estimated fundamental frequency ω values accurately reflect physiological heart rates across time, even as they change with drug dose. Further, due to the observations regarding time-frequency resolution above, the algorithm presented here can provide far greater precision than common clinical measures and thus enable accurate estimation of the parameters and the BCG artifacts continuously over time.

Finally, the robust and accurate convergence of the 1-D optimization for ω expressly implies (by separation theorems detailed in [156]) that the other parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and σ^2 also converge to their global optima. Having thus established the accuracy and physiological relevance of our parameter estimates, we proceed to review the spectral and time domain estimates of the BCG artifact and clean EEG for our test datasets.

2.5.3 Recovery of Oscillatory Dynamics

Fig. 2-6 and Fig. 2-7 show the raw EEG measurement, estimated BCG artifacts and clean EEG for oscillatory dynamics from the propofol-induced anesthesia paradigm in § 2.4.1.1. Harmonic regression estimates, and a comparison with the motion-sensor based adaptive filtering are shown.

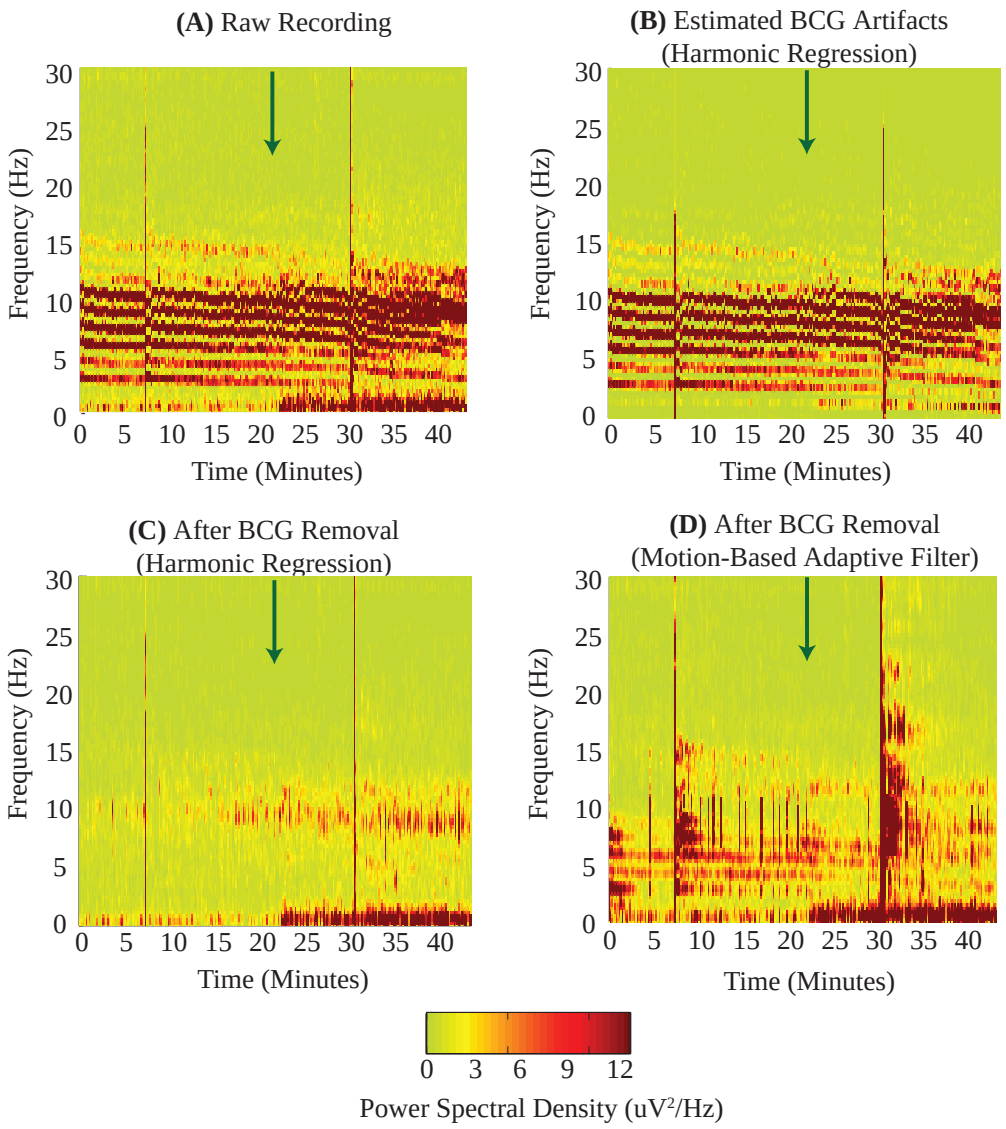

Figure 2-6: Harmonic Regression Algorithm Effectively Removes BCG Artifacts from Oscillatory Brain Dynamics Recorded at 3 T - Spectral Domain Results. Estimates from the algorithm are shown in spectral domain for a representative 45 minute segment of frontal EEG recordings acquired during propofol anesthesia in the scanner. Green arrows indicate time of increase in propofol dose, and mark time points when oscillatory changes of interest are expected to appear in the EEG. *(A-B)* Spectrograms of raw BCG-corrupted EEG recording and estimated harmonic BCG artifacts respectively. *(A vs. B)* illustrates that the algorithm effectively removes harmonic BCG artifacts. *(C)* Spectrogram of the clean EEG estimates (AR term) shows oscillatory power in the $0-1$ Hz slow-band and in the $8-12$ Hz alpha-band, as expected during propofol anesthesia. Comparable oscillations in ground truth recordings obtained under similar conditions outside the scanner for a different subject are in Fig. 2-2*A*. Thus, our algorithm indeed restores true drug-induced oscillatory EEG signatures of interest. *(D)* Spectrograms of clean EEG estimates recovered by an motion-sensor based adaptive filtering approach [100] implemented as in § 2.4.2.5. *(C vs. D)* Contrasting the efficacy of the two algorithms, it is seen that, when using the motion based adaptive filter, the drug-induced EEG oscillatory features of interest are not as evident, and large residual artifacts are left behind. Gross motion artifacts manifest in the spectra as high power across frequencies (vertical lines).

Fig. 2-6 shows spectral domain estimates. Inspecting the raw measurement ($\mathbf{y}$) and harmonic BCG estimates ($\hat{\mathbf{s}}$) in Fig. 2-6*A*-*B*, we note that the algorithm effectively takes out the harmonic artifacts. The clean EEG estimate ($\hat{\mathbf{v}}$) after BCG removal is in Fig. 2-6*C*, revealing oscillatory power in the $0-1\,\text{Hz}$ slow-band and $8-12\,\text{Hz}$ alpha-band at the expected time (green arrow marking propofol dose increase). These slow and alpha oscillations are also seen in ground truth spectra of EEG data obtained outside the scanner under a similar propofol anesthesia protocol for a different subject (Fig. 2-2*A*, [150]). Thus, our model-based approach and algorithm can effectively distinguish overlapping bands of BCG artifact (harmonic streaks) and brain-generated EEG activity (oscillatory features) at 3 T.

Fig. 2-6*D* shows, for comparison, the clean EEG estimates obtained with motion-sensor based adaptive filtering for the same data. While the adaptive filter can remove much of the harmonic streaks, it leaves behind significant residuals, does not reveal the increased $8-12\,\text{Hz}$ alpha-band activity, and is also more sensitive to other EEG artifacts such as the gross motion artifacts (seen as vertical lines).

Fig. 2-7 shows the time domain estimates. Inspecting the harmonic regression estimates in Fig. 2-7*A*, we see that the our algorithm effectively takes out harmonic combs $\hat{\mathbf{s}}$ from the raw data $\mathbf{y}$ (top panel). The harmonic combs comprise a large portion of the raw data power. The clean EEG estimates $\hat{\mathbf{v}}$ that are left behind (middle panel) contain lower amplitude EEG oscillations (Fig. 2-7*A*). Specifically, a slow oscillatory envelope with ongoing alpha activity atop it is seen. The residual series $\hat{\boldsymbol{\epsilon}}$ (bottom panel) contains no significant temporal structure (corresponding goodness of fit analysis shown in Fig. 2-4*D*-*E*).

Fig. 2-7*B* shows, for comparison, the motion-sensor based adaptive filter results. Inspecting the BCG artifacts estimated by this algorithm, we see that they do not fit the harmonic combs in the raw data as well (top panel). Thus, clean EEG estimates (middle panel) contain residual artifacts ($0-1.5\,\text{second}$), and additional high amplitude high frequency features not present in the raw data ($1.5-2.5\,\text{second}$). Thus, it appears that the algorithm sometimes introduces artifacts. To understand why, we inspected the motion sensor recordings that were used as a reference template to inform the adaptive filter (bottom panel). The reference motion sensor signals are particularly noisy in the $1.5-2.5\,\text{second}$ range, thus corrupting the estimates. Our algorithm does not face issues arising from corrupted reference recordings as it is reference-free and relies solely on the native structure in the BCG artifacts.

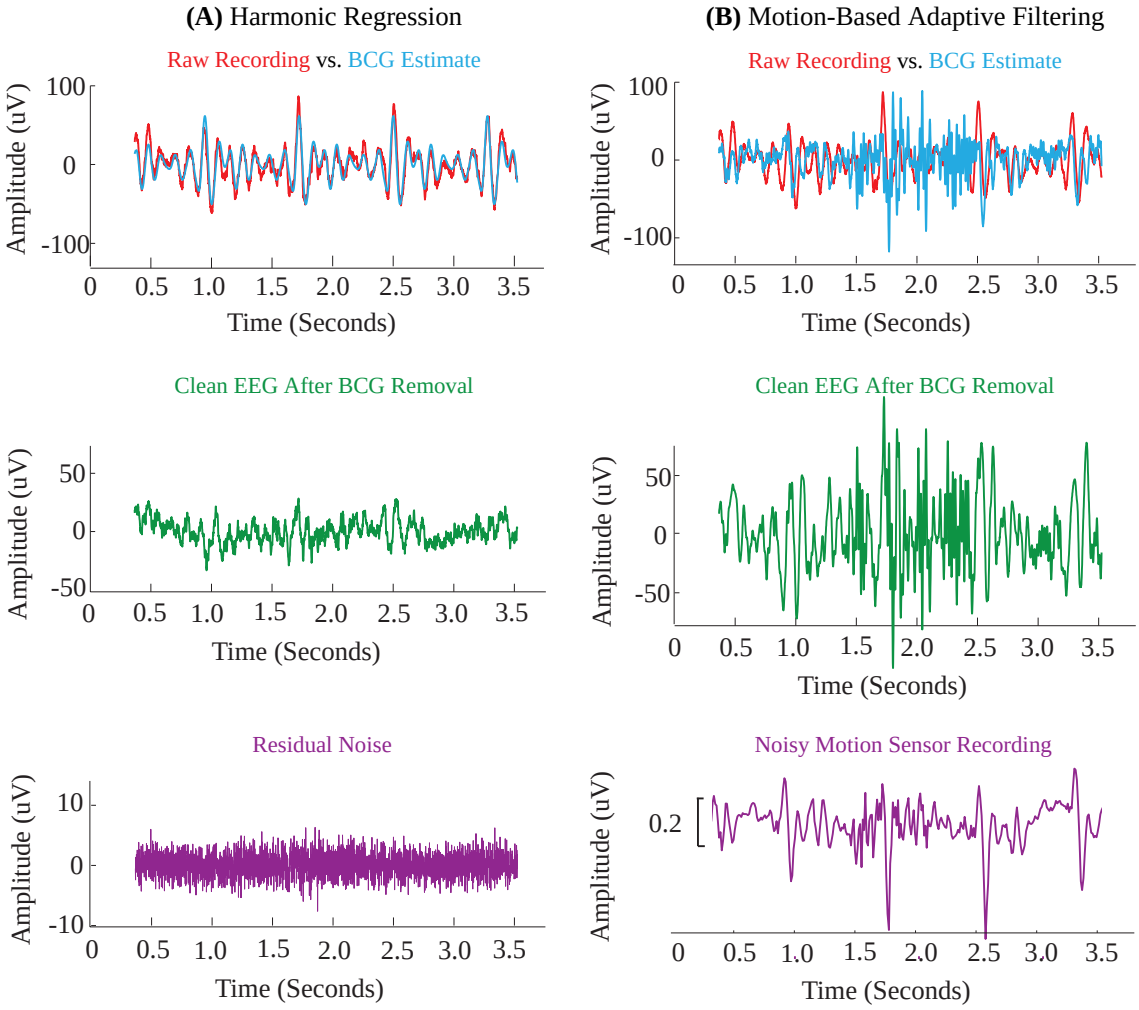

Figure 2-7: Harmonic Regression Algorithm Effectively Removes BCG Artifacts from Oscillatory Brain Activity Recorded at 3 T - Time Domain Results. Estimates from the harmonic regression and motion-based adaptive filtering algorithms are shown in time domain for a representative 3 second segment of frontal EEG recordings within the propofol anesthesia data in Fig. 2-6. In each case, the raw data series is in red, the estimated BCG artifacts overlay is in blue, and clean EEG estimates (after subtracting BCG from raw data) are in green. *(A)* The harmonic regression algorithm effectively removes BCG artifact combs from the raw data series (top panel), reveals clean EEG estimates (middle panel) with ongoing slow-band and low amplitude alpha-band oscillations, and leaves a residual with no significant temporal structure (bottom panel). *(B)* The motion-based adaptive filter algorithm does not fully remove BCG artifact combs from the raw data (top panel) and reveals EEG estimates with large residuals and additional artifactual features which obscure oscillatory features of interest (middle panel). These residuals and artifacts are introduced by noise inherent in the motion sensors used as a template for adaptive filtering (bottom panel).

2.5.4 Recovery of Evoked Responses

Fig. 2-8*A*-*C* show time domain estimates for evoked responses recorded per the visual checkerboard paradigm described in § 2.4.1.2. Harmonic regression estimates, and a comparison with the ECG-

based subtraction algorithm are shown.

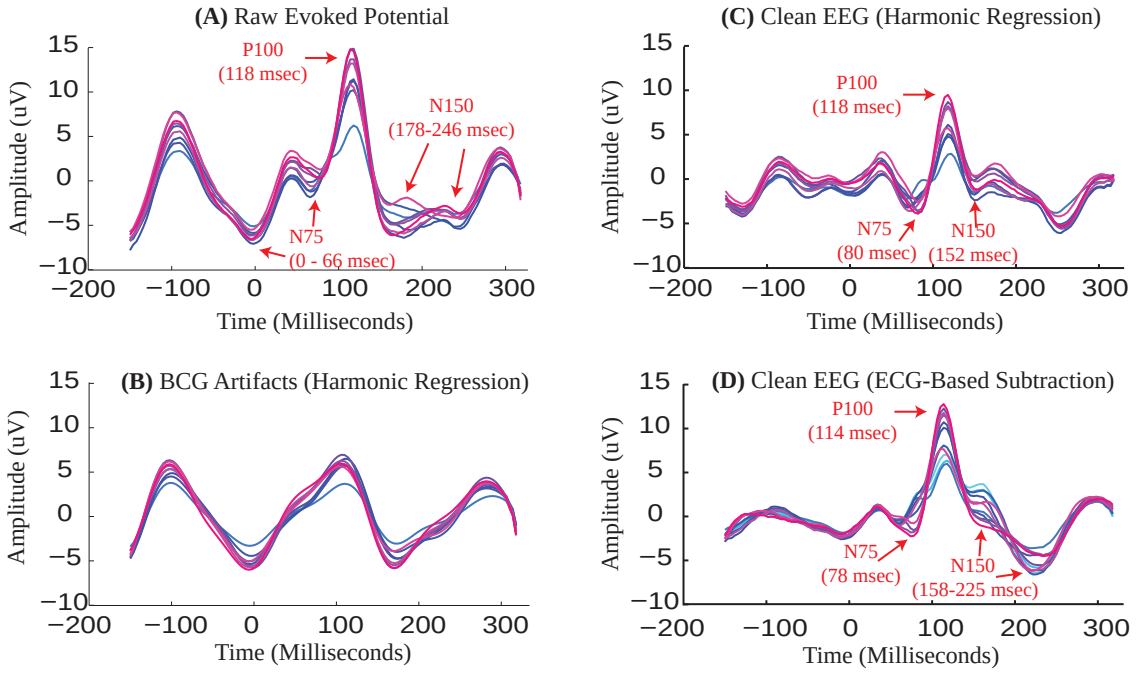

Figure 2-8: Harmonic Regression Algorithm Effectively Removes BCG Artifacts from Evoked Responses Recorded at 3 T. All evoked responses are presented as stimulus-locked averages across 9 minutes of recordings in right occipital channels (each color corresponds to one occipital channel). Panels *(A-C)* show raw BCG-corrupted evoked response, harmonic regression estimate of BCG artifacts, and clean evoked response estimates respectively. *(A vs. B)* While the P100 peak is visible in the raw recording, the amplitudes and latencies of the N75 and N150 peaks appear unclear (possible ranges that can be inferred from the peaks in the data are shown in red). Further peaks around −100 msec and 300 msec distract from the shape of the raw evoked response. *(B-C)* Harmonic regression removes the periodic BCG artifacts, and reveals clean EEG with N75 and N150 peaks seen more clearly at expected latencies [130, 159]. These results show that the proposed harmonic regression algorithm effectively uncovers evoked EEG features of interest and generalize efficacy of BCG removal across varied EEG-fMRI experimental paradigms. *(D)* Clean evoked responses estimated from the same data using an ECG-based subtraction algorithm [80] implemented as in § 2.4.2.5. *(C vs. D)* While the subtraction algorithm also cleans much of the BCG artifact and restores the N75 peak, the cleaned evoked responses have larger amplitudes and the N150 peak is less clear.

The visual evoked paradigm is known to typically elicit three peaks in normal subjects [130, 159]: namely the N75 with onset latency $75-80$ msec, P100 with onset latency $95-115$ msec and N150 with onset latency $150-170$ msec. However, in the scanner, while the raw EEG measurement in the scanner (Fig. 2-8*A*) shows the P100 peak, the location and latencies of the N75 and N150 peaks appear unclear and/or aphysiologic (earlier or later than expected). The harmonic regression estimate of the BCG artifact (Fig. 2-8*B*) comprises multiple harmonics of the fundamental frequency, which, when averaged across epochs, appears as a large periodic component with a mean peridicity of ~ 5 Hz. The clean EEG estimate after BCG removal (Fig. 2-8*C*) preserves the P100 peak, but also clarifies the N75 and N150 peaks at the expected latencies. Thus, our algorithm is capable of effectively removing BCG artifact and recovering expected evoked potential signatures.

Fig. 2-8*D* shows, for comparison, the clean EEG estimates obtained with ECG-based subtraction for the same data. While the ECG-based method preserves the P100 peak, and clarifies the N75 peak to expected latencies, it is less clear whether there is an N150 peak and at what latency it occurs. As ECG-based methods subtract out the mean of data windows time-locked to the heart beat, these methods likely subtract out temporally overlapping features of the BCG artifact and the true ERP thus leading to missed or delayed peaks. Further, we note that amplitudes of the ECG-based estimates are significantly higher than the harmonic regression estimates, suggesting that the former may have left behind more residual artifact than our algorithm.

2.5.5 Validation Summary

In summary, we have shown that (a) the model fits the form of the data well, (b) the algorithm estimates parameters robustly, efficiently and accurately and (c) the procedure for obtaining parameter estimates is well-suited to temporal and spatial variations in the BCG artifacts. Estimation results on two distinct paradigms of artifact corrupted data illustrate that the algorithm can remove artifacts effectively and restore signatures of interest. Although BCG artifacts tend to be more problematic in fronto-lateral EEG channels where movement cannot be effectively restrained, than in occipital channels, our model-based technique preserves temporal structure of the true brain-generated EEG just as well in frontal channels (Fig. 2-6-Fig. 2-7) as it does in occipital channels (Fig. 2-8). Next, we discuss simulation tests to compare performance and recovery of signals both with our algorithm and existing reference-based algorithms.

2.6 Performance Analysis

2.6.1 Simulation Tests for Benchmarking

To test algorithm performance with simulated signals, it is important to have realistic artifacts. Thus, resting state EEG data were collected on 2 healthy volunteers in a static 3T magnetic field (Siemens, Erlangen, Germany). Study subjects were asked to lay awake and motionless (with eyes open) inside the scanner for 5 minutes. As the brain generated EEG in this paradigm contains few structured time-frequency EEG features, the recorded EEG signal predominantly contains the BCG artifact, providing realistic artifacts for testing with simulated signals. Continuous EEG was acquired at 957 Hz using an MR-compatible low noise, high dynamic range 24-bit electrophysiological recording system [94]. EEG electrodes made of Ag/Ag-Cl were placed in adjacent bipolar pairs in 8 locations (M2 →T8, T8→C6, C6 →C4, C4→Cz, Cz→C3, C3→C5, C5→T7, T7→M1) across the central coronal plane. ECG data were simultaneously acquired with the same amplification and acquisition system as the EEG for comparisons between our reference-free approach and reference-signal based BCG removal techniques. This study was approved by the Partners Human Research Committee at Massachusetts General Hospital, and all subjects provided informed written consent.

For each subject, we simulated 35 oscillatory EEG test cases with varying amplitudes (9 μV, 15 μV, 21 μV and 30 μV) for a range of SNRs, and varying frequency content (across delta, theta, alpha and beta bands), similar to those seen in sleep, attention, memory and anesthesia, to benchmark performance of our algorithm. In each test case, the simulated signature was added on to the BCG artifact recording to produce a BCG-corrupted EEG oscillation for assessing performance of our algorithm. Fig. 2-9 illustrates the design of these test cases.

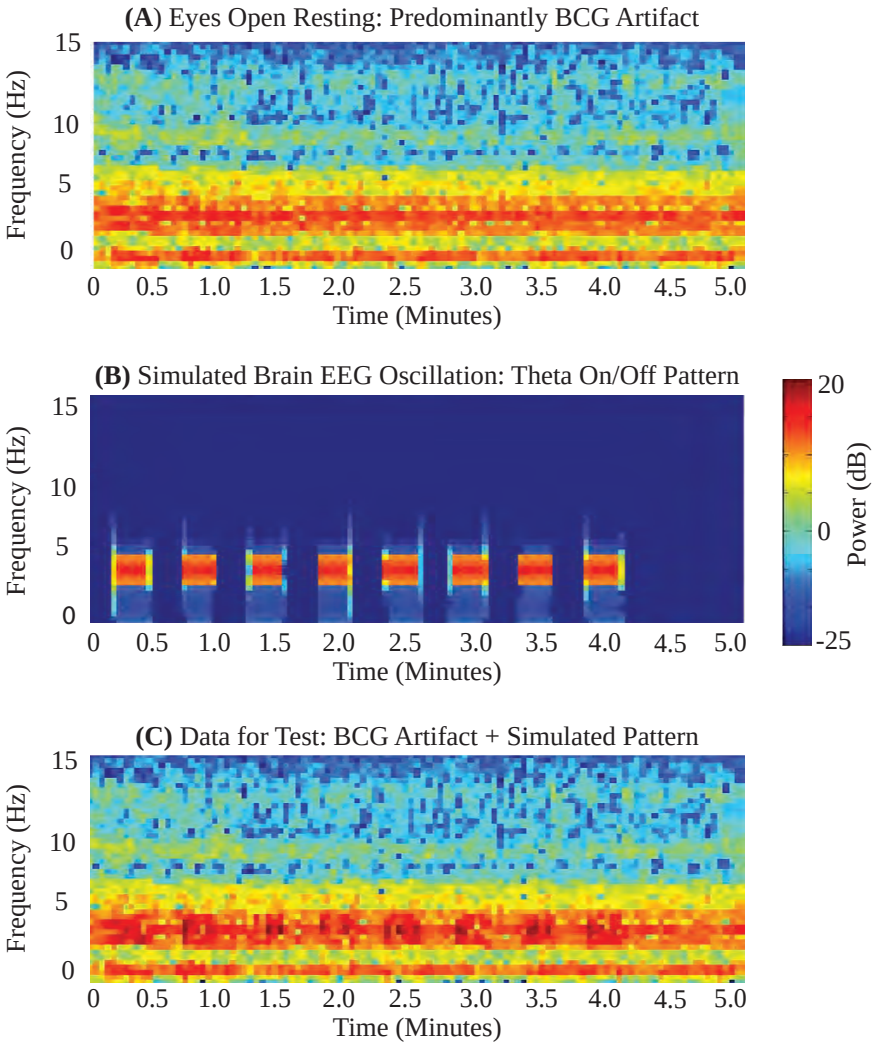

Figure 2-9: Simulation Tests for Benchmarking. *(A)* Example eyes open resting state EEG recording largely comprising the BCG artifact. *(B)* A $3-4$ Hz test theta oscillation with a periodic 17 second ON/17 second OFF pattern, and ON period amplitude of 21 μV. *(C)* The simulated test oscillation is added to resting state EEG recorded in the MRI scanner to produce BCG-corrupted EEG oscillations, which are used to assess performance of artifact removal algorithms. All plots are for a temporal channel (M2 $\rightarrow$T8).

With these test cases, we assessed performance of the harmonic regression algorithm in relation to that of the commonly used ECG-based average subtraction and optimal basis set (OBS) algorithms. First, we compared spectral and time domain results with the ground truth simula-

tions. This enables us to study whether the BCG removal introduces any systematic bias in the cleaned EEG estimates. Knowing whether a BCG removal algorithm, in removing artifacts, changes anything about the underlying signal (especially power and phase) is relevant for neuroscience applications. Fig. 2-10 shows comparisons between clean EEG estimates after BCG removal and the simulated signals on a temporal channel. The spectrograms (Fig. 2-10*A*-*B*) show that harmonic regression enables improved recovery of power and timing of the simulated signal, in comparison with OBS. The time series plots (Fig. 2-10*C*-*D*) show that harmonic regression enables better recovery of phase and amplitude of the simulated oscillation than does OBS.

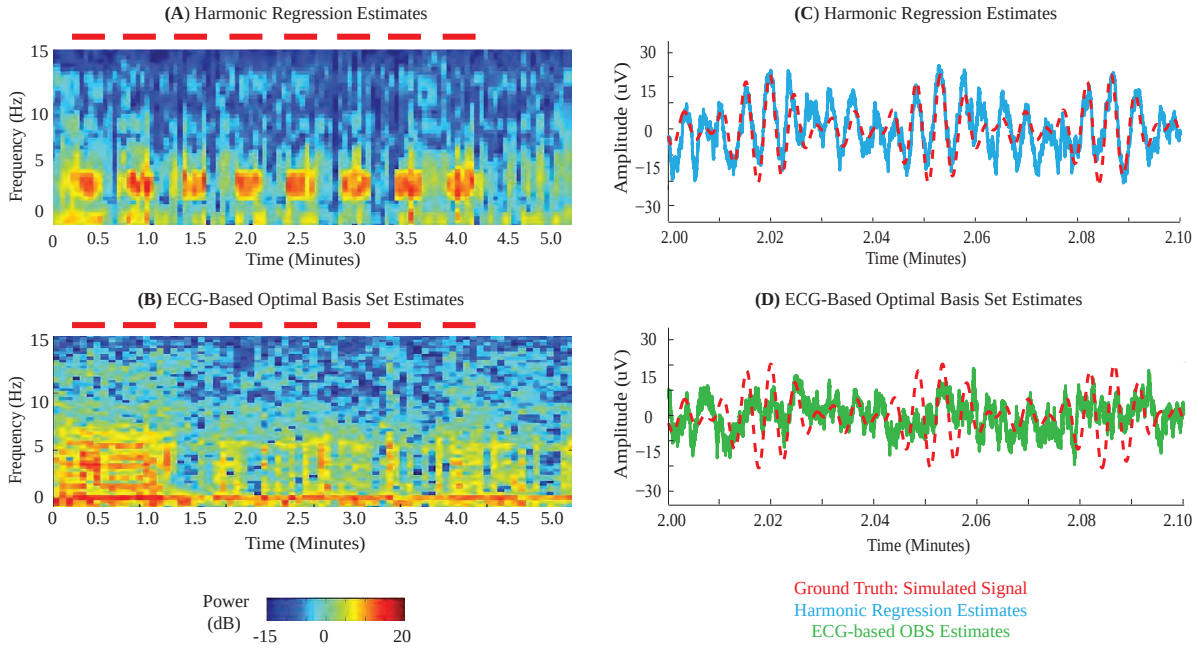

Figure 2-10: Comparison of Estimates with Ground Truth Simulations. *(A-B)* Spectrograms of clean EEG estimates after BCG removal using harmonic regression and OBS. Red dashes mark times at which the $3-4$ Hz simulated activity is ON. While the simulated pattern, timing and amplitude are clearly restored in the harmonic regression case, recovery is less apparent with OBS. *(C-D)* Time series estimates from harmonic regression and OBS, overlaying the simulated oscillatory signal. The harmonic regression estimates match the amplitude and phase of the simulated oscillation, while OBS removes signal in addition to artifact and thus, does not restore the simulated activity. All plots use same data, temporal channel and color scale as shown in Fig. 2-9.

Second, we quantified power SNR improvement in the frequency domain using a method proposed in [93]. As the test case has a simulated signal that has known ON and OFF periods, it is possible to quantify the SNR in the frequency band of interest before and after artifact removal. Then, the SNR improvement is defined as $\frac{\text{SNR}_{\text{raw}}}{\text{SNR}_{\text{clean}}} - 1$, and averaged across all ON/OFF periods within the test. Table 2.1 shows the SNR improvement values for the different test cases and algorithms. It is noteworthy that the harmonic regression algorithm performs just as well for low amplitude signals (low SNR) as it does for larger signals (higher SNR).

Delta-Theta Oscillation Amplitude	Harmonic Regression	ECG-Based Average Subtraction	ECG-Based Optimal Basis Set
$9\,\mu$V	5.3	1.0	4.6
$15\,\mu$V	6.9	1.8	1.1
$21\,\mu$V	7.2	2.1	1.0
$30\,\mu$V	7.0	2.2	1.1

Table 2.1: SNR Improvement for Different Test Cases and Algorithms. The harmonic regression algorithm consistently outperforms reference-based methods, even when the simulated signals have low $9\,\mu V$ amplitude. ECG-based average subtraction appears to perform better than OBS when test oscillations have significant amplitudes. However, while quantitative performance of the ECG-based algorithms is better than OBS, the general trends, in recovering timing, phase and amplitude of the simulated signals, seen in Fig. 2-10 for OBS hold for ECG-based average subtraction too.

2.6.2 Analysis of Improved Performance of Harmonic Regression

To better understand our results, we now analyze performance of the reference-based and harmonic regression algorithms in relation to each other. First, ECG-based methods have been shown to work reasonably reliably in cases where evoked signatures are being analyzed. However, our simulation tests analyzed oscillatory activity. ECG-based methods essentially estimate the BCG as the "average" shape of the EEG within a window around the ECG R-wave (where "average" can be derived via moving average, principal component analysis or other methods). However, when ongoing brain oscillatory activity overlaps this window in time, the brain oscillatory signals too can get marked as artifacts. Thus, these methods have the tendency to discard true EEG. Further, we note that the ECG-based methods can leave large residuals by ignoring artifact features outside these windows. Fig. 2-11*A* illustrates this point by showing that the BCG artifacts (estimated by our harmonic regression algorithm) are high even at times outside of the artifactual window considered by ECG-based algorithms. While several efforts have been focused on (a) overcoming ECG peak detection problems and (b) developing better ways to identify the "average" shape of the EEG within the windows around the heart rate, our analysis suggests that the limitation in the ECG-based methods is more fundamental, and related to the way the BCG artifact is defined as a discrete occurrence within a predefined time window around the ECG R-wave.

Second, while Fig. 2-7 showed that the performance of motion-based adaptive filtering suffers when the reference motion recordings are noisy, we further analyzed the relation between the motion signals and the EEG data. Specifically, as coherence between the motion signal and the EEG data indicates frequencies where the adaptive filtering-based BCG estimates are concentrated, we plotted coherence between the motion signal and EEG data (Fig. 2-11*B*). We note that the coherence between EEG and motion signals is not limited to harmonics but includes band-specific activity. Thus, the motion-based adaptive filter may estimate the BCG as a non-harmonic band-

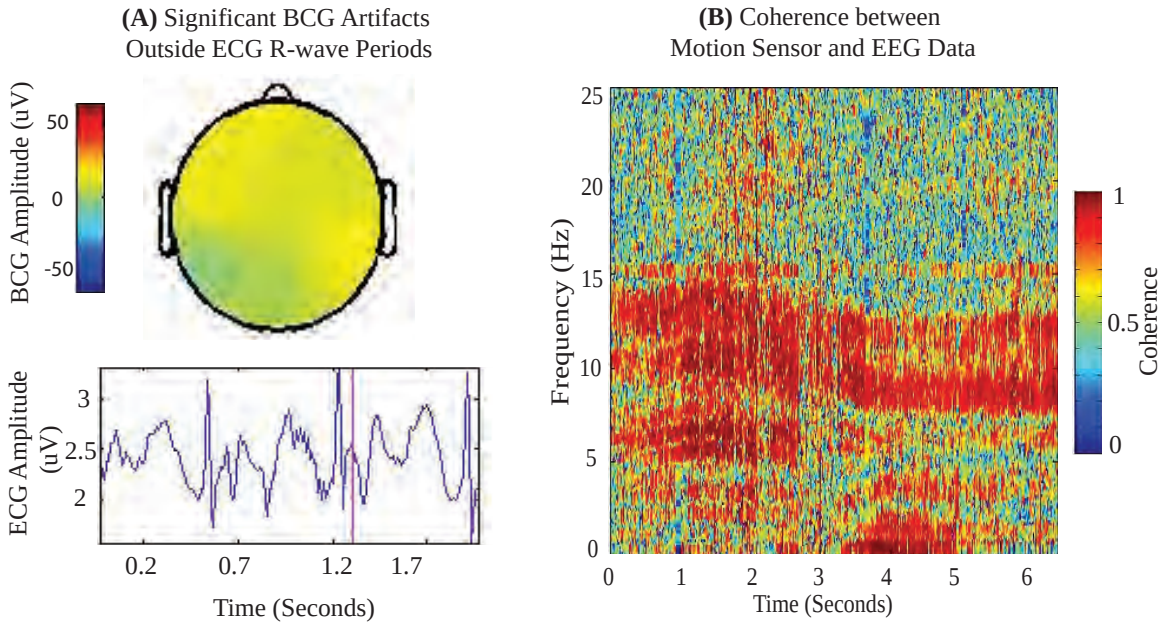

Figure 2-11: Models Implicit in Reference-Based Approaches Affect their Performance. *(A)* Top topographical plot shows harmonic regression estimates of the BCG artifact, across channels, at a time point within the heart cycle indicated by the pink line on the bottom ECG plot. ECG-based subtraction approaches treat the BCG as a discrete occurrence in a predefined artifactual window around the R-wave of the ECG. But the time point marked, outside the usual predefined artifactual window, has BCG artifacts with significant amplitude ranging $5 - 25\,\mu$V. ECG-based approaches neglect artifacts outside of the predefined window, thus leaving behind large residual artifacts. *(B)* Coherogram between a motion recording and EEG data from the propofol oscillatory study (§ 2.5.3). Motion-based adaptive filtering treats the BCG artifact as a linear function of the motion sensor recording. However, in cases when coherence between the motion recording and EEG data is not harmonic, such as in the case shown, the BCG artifact estimates may not have the inherent harmonic structure and thus residual artifacts may be left behind.

specific process, to leave behind residuals of the BCG harmonics, and/or discard true band-specific EEG activity. We note that this adaptive filter performs better in cases when the motion-EEG coherogram has harmonic structure than in cases when it does not.

Overall, the above analysis suggests that reference-based methods are not only limited by noise corrupting the reference sensors, but also by the fact that the models they use (implicitly or explicitly) do not always reflect the strong harmonic structure and temporal features of the BCG artifact. Thus, the physical and physiologic correspondence of our model to the nature of the BCG artifacts is important for the improved performance we are able to achieve. That said, while our model-based procedure does not require reference signals, it is possible that incorporating specific information from the reference sensors within our approach could provide additional information. For example, the ECG could be used to further inform the range of possible fundamental frequencies within our algorithm. Further, the motion sensors could be used to inform occurrence and removal of gross motion artifacts that are both larger and distinct from the BCG artifact.

2.6.3 Remarks on Sensitivity to Analysis Choices

The main parameters used in our algorithm and analysis are the model orders and the length of the moving window. While the performance of the harmonic regression algorithm is relatively insensitive to the model orders within a reasonable range (plots of information criteria in Fig. 2-4), it is sensitive to the length of the moving window T. Specifically, as the algorithm fits only one set of parameters for a given window of data, the amplitude of the harmonic is maintained uniform throughout the segment. Hence in cases where the BCG amplitude varies widely between consecutive heart beats, the performance of the algorithm is sensitive to choice of window length. An example of this problem is illustrated in Fig. 2-12. Regardless of this sensitivity, the algorithm can cope with variation if moving window length is chosen suitably. Further, the parameter estimates inherently quantify the variations in artifacts over time. It is also possible to overcome this issue by allowing the window to have variable length when large variations in heart beat occur (such as with change in drug dose or physiologic status) rather than maintaining always the same length T. Additionally, extensions to recursive state space methods where the estimation is done recursively with each new datapoint will overcome this issue, as the parameters can then evolve instantaneously with the data.

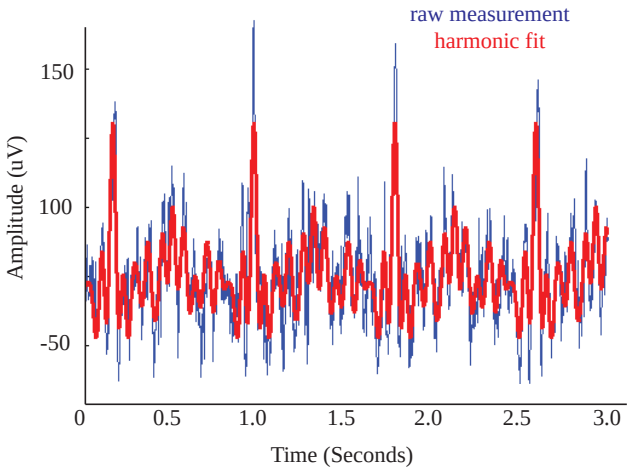

Figure 2-12: Sensitivity to Moving Window Length. The harmonic fit to the raw measurement over a 3 second segment of data corrupted with artifacts. As the algorithm fits only one set of parameters for a given window of data, the amplitude of the harmonic is maintained uniform throughout the segment. However, the underlying BCG varies widely between consecutive heart beats. While this is an extreme example of beat-to-beat variation in BCG amplitude, in such cases the length of the moving window should be shortened to about $1 - 1.5$ second for good performance of the algorithm.

2.7 Discussions

EEG-fMRI offers uniquely high spatiotemporal resolution for functional brain imaging, but its widespread adoption and utility for neurophysiologic studies has been limited due to challenging

ballistocardiogram (BCG) artifacts. We have developed a model-based, reference-free and statistically principled algorithm to remove these artifacts. Our approach (a) exploits the structure in the physics and physiology underlying these artifacts to define the BCG as a harmonic process, (b) frames the BCG removal problem as a model-based parametric estimation problem and (c) adapts a local likelihood estimation algorithm to solve this non-linear time-varying parameter estimation problem and remove the artifacts. We have shown that (a) the model fits the form of the data well and (b) the algorithm enjoys unique optimality and convergence properties. Further, we have performed extensive validation tests to demonstrate that the algorithm (a) effectively removes artifacts from EEG recorded in the MRI scanner, (b) restores ground truth oscillatory and evoked EEG signatures, including specific timing and power changes expected from similar experiments performed outside the scanner, and (c) affords improvements over existing reference-signal based BCG removal techniques. These improvements are enabled by the unique features of our model-based statistical estimation approach, as detailed below.

2.7.1 A Model-Based Approach for BCG Artifact Removal

The BCG removal problem presents three concurrent challenges of high amplitude artifacts, time-frequency overlap, spatiotemporal variations - our approach offers advantages with respect to each.

First, our approach is well-suited to cope with high amplitude artifacts. This is because it exploits the high artifact to signal ratio in the problem by targeting the clear, physically based harmonic structure in the artifacts. Given the apparent form of the artifact, reference-signal based templates for the artifacts [80, 100, 138, 139] or component analysis techniques [101, 102, 143, 145] that empirically derive the basis for the artifact are not required. In fact, our approach removes artifacts more effectively (less residual power) than methods that rely on (a) noisy reference signal-based templates, (b) pre-defined time intervals for when the BCG occurs (as artifacts are significant even outside of these intervals), or (c) component separation criteria that do not fully decompose the raw EEG into artifact vs. signal.

Second, our model is well-suited to cope with time-frequency overlap between artifact and signal. This is because it includes a loose autoregressive term that can take on the temporal structure of the neurophysiologic EEG data. This enables our approach to preserve brain-generated EEG signatures of interest. While many existing methods are able to remove a good proportion of the BCG artifacts, often they are unable to retain or recover signal features like evoked peaks, oscillatory power underlying harmonic bands, or phase-amplitude coupling between bands that overlap with the artifact in time or frequency domains [95, 96].

Third, our algorithm copes well with spatiotemporal variations in the BCG artifact, and provides accurate estimates of brain-generated EEG signals as they evolve. This is because it is framed as a parametric model-based time series estimation problem which lends itself well to time-varying parameter estimation techniques, model selection procedures and goodness of fit analyses. Specifi-

cally, our local likelihood-based estimation allows the estimates to adapt to temporal variations in both the BCG artifact and neurophysiologic EEG signals. Further, by independently performing the estimation for each channel, we are able to cope with spatial variations in the BCG.

Much of the BCG removal literature has used ready black-box approaches which (a) inherently employ (hidden) models that are not customized to the structure of the BCG artifact problem, (b) rely on noisy reference data templates, and (c) do not offer clear statistical characterizations to explain the data. Our approach and results motivate the idea that employing explicitly stated models customized to the inherent nature of the data, and developing statistical methods that use these models to explain the data can be more effective for this challenging problem.

2.7.2 Practicability for a Variety of Imaging Paradigms

Our algorithm is practicable in a variety of imaging and acquisition conditions. First, there is increasing interest in acquiring EEG-fMRI at high fields ($> 3\,\mathrm{T}$) due to the improved spatial resolution and higher functional sensitivity of fMRI at these field strengths [160, 161]. While EEG-fMRI acquisition systems have been set up for high fields [94, 162], these systems have limited utility at present as BCG artifacts corrupting the EEG become more severe at higher static fields [80, 93, 136, 137]. As our approach builds on the high artifact to signal ratio within the BCG removal problem, and has demonstrated utility in large artifact low SNR conditions such as frontolateral channels and small amplitude oscillations, it is well-poised for these high field studies. Second, neurophysiologic and clinical questions often necessitate interleaved EEG-fMRI scanning - where short segments of continuous EEG data are interspersed with large EPI artifacts during fMRI acquisition. The sliding window approach makes our technique just as well-suited to the short lengths of EEG data recorded in interleaved scans, as for longer segments of data in continuous EEG-fMRI. This is distinct from subtraction or basis decomposition techniques [80, 101, 102], where long segments of continuous EEG data are required for estimation. Finally, our method is amenable to real-time implementation – and is thus suited for continuous monitoring and online artifact removal. All these features suggest that our algorithm is practical and generalizable to a range of study protocols, and can be incorporated within EEG-fMRI acquisition systems for robust, sensitive, and specific EEG-fMRI measurements.

2.7.3 Implications for Neuroscience Studies

The vast majority of simultaneous EEG-fMRI studies reported in the neuroscience literature have focused on regional correlates of EEG time-frequency features recorded in (a) occipital channels (e.g., eyes closed alpha oscillations or visual evoked responses [64, 71, 72]), (b) high frequency bands (e.g., γ-band oscillations or the $40\,\mathrm{Hz}$ auditory steady state response [93, 163, 164]), and (c) short duration studies spanning a few minutes. Coincidentally, all these cases represent less troublesome BCG artifacts [80, 100] as: (a) pulsatile head movement can be constrained more effectively for

occipital channels (as opposed to frontolateral channels), (b) BCG artifacts typically do not much affect high frequencies as they primarily corrupt $0-25\,\text{Hz}$ frequencies, and (c) BCG time variations are less severe for short few-minute periods (as opposed to long studies). Further, such test cases and characterizations have also been the focus of BCG removal literature [80, 95, 96, 100, 102, 139, 140], perhaps, due to the fact that recordings of visual activity or short durations are easily accessible test cases.

However, EEG signatures in frontal, central and temporal channels and $0-25\,\text{Hz}$ frequencies evolving over long duration recordings are of great interest for applications in studies of memory, attention, decision making, awareness, sleep, anesthesia, and epilepsy (§ 1.2). We have tried to include testing and validation in cases relevant to these potential application areas. Our validation includes frontal channels that are known to have severe artifacts, as well as quantitative benchmarking in recordings on central and temporal channels. Further, we show recovery of slow, theta and alpha bands ranging $0-15\,\text{Hz}$. Finally, we demonstrated performance in a drug-induced setting with large time-variations (in heart rate and thus fundamental frequency) over close to an hour. These results establish that our method is uniquely suited to application areas of wide relevance in neuroscience investigations.

Further, good performance across channels affords a clear advantage for EEG-fMRI applications in electrical source imaging and coherence or connectivity studies where clean EEG across the head is desirable [104, 165, 166]. Finally, we note that our parametric model-based approach has the advantage that it can help probe a variety of specific EEG signatures of interest in a statistically principled manner – as the time series analysis can be framed to explicitly test hypotheses relating to neurophysiologic EEG changes arising as a result of stimulus, drug or other experimental intervention.

2.7.4 Future Directions

Future investigations will study recursive or state space implementations of our estimation procedure for the harmonic regression in colored noise problem. This will overcome the sensitivity to moving window length and enable greater ability to cope with beat-to-beat variations in the BCG artifact while taking continuous evolution of the neurophysiologic EEG state into consideration. Further refined understanding of the various sources of the BCG artifact and their respective contributions to this artifact can improve the proposed model, and incorporation of better models for EEG evoked peaks and spikes can generalize the approach further. Finally, further evaluations on continuous EEG-fMRI scans with different imaging protocols, and on other quantitative EEG features of interest (e.g., K complexes, epileptic spikes, sleep stage scoring, or oscillatory coupling) will validate the algorithm for areas of interest in cognitive or clinical neurophysiology studies.

2.8 Appendix

2.8.1 Prior Covariance on Model Parameters

Given the time-frequency overlap in this problem, the estimates of the harmonic amplitudes $\boldsymbol{\beta}$ would be best informed by the AR spectral structure specified by AR covariance $\mathbf{Q}$. However, as the algorithm does not initially have an estimate of $\mathbf{Q}$ at the start of cyclic descent (we initialize with $\mathbf{Q} = \mathbf{I}_{T\times T}$), the estimates of the harmonic amplitudes $\boldsymbol{\beta}$ need to be informed with prior information while the cyclic descent proceeds and estimates $\mathbf{Q}$ from the data. For this purpose we derive a prior covariance matrix $\mathbf{W}$ for $\boldsymbol{\beta}$, as detailed below.

For the generalized least squares problem denoted by $\mathbf{y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{v}$ where $\mathbf{Z}$ is a known harmonic basis, $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ and $\boldsymbol{\beta}$ is drawn from a distribution with prior covariance $\mathbf{W}$, we can write the negative log-likelihood as:

$$-\log\, L_{GLS} = -\frac{1}{2}(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})'\mathbf{Q}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta}) - \frac{1}{2}\boldsymbol{\beta}'\mathbf{W}^{-1}\boldsymbol{\beta}. \tag{2.25}$$

The optimal solution for $\boldsymbol{\beta}$ is given by minimizing $-log\, L_{GLS}$ as below:

$$\frac{d\,(-\log\, L_{GLS})}{d\boldsymbol{\beta}} = \frac{1}{2}\frac{d}{d\boldsymbol{\beta}}(-\mathbf{y}'\mathbf{Q}^{-1}\mathbf{y} + \boldsymbol{\beta}'\mathbf{Z}'\mathbf{Q}^{-1}\mathbf{y} \tag{2.26}$$

$$+ \mathbf{y}'\mathbf{Q}^{-1}\mathbf{Z}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{Z}'\mathbf{Q}^{-1}\mathbf{Z}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{W}^{-1}\boldsymbol{\beta}) = 0. \tag{2.27}$$

Simplifying this, we have:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{Z}'\mathbf{Q}^{-1}\mathbf{Z} + \mathbf{W}^{-1}\right)^{-1}\mathbf{Z}'\mathbf{Q}^{-1}\mathbf{y}. \tag{2.28}$$

Viewed in this way, the prior covariance $\mathbf{W}$ and the AR term covariance $\mathbf{Q}$ determine what proportion of total power in the measurement $\mathbf{y}$ is allocated to harmonic BCG artifact versus to the brain-generated AR EEG estimate. If $|\mathbf{W}| >> |\mathbf{Q}|$, then nearly all measurement power in the harmonic bands is allocated to the harmonic BCG artifact, but if the AR structure $\mathbf{Q}$ needs to inform the harmonic fit and preserve EEG structure, lower values of $|\mathbf{W}|$ are desirable. Assuming the prior covariance $\mathbf{W} = E(\boldsymbol{\beta}'\boldsymbol{\beta})$ is diagonal and can be treated as the desired harmonic power, we specify the j^{th}diagonal term of $\mathbf{W}$ as the difference between the power in the j^{th} harmonic line and the background AR power in a 2 Hz band centered around the j^{th} harmonic line. Thus, when the AR power in the neighborhood of the harmonic streaks is high, the $\mathbf{W}$ discounts the harmonic amplitude to retain some of the underlying EEG power.

For each possible value of fundamental frequency ω, we pre-compute the corresponding $\mathbf{Z}$ and

the prior covariance $\mathbf{W}$, so the appropriate $\mathbf{W}$ is used for a given guess of the fundamental frequency during the nested cyclic descent iterations. In practice, for our nested cyclic descents, we note that the prior covariance primarily affects the estimated $\boldsymbol{\beta}$ and does not change the ω estimate much. This formulation also enables formal regularization of solutions as needed.

2.8.2 Implementation Details

Test Case	**Number of Harmonics** R	**Number of Autoregressive Terms** P	**Moving Window Length** T
Propofol Oscillations	16	8	3 seconds
Evoked Responses	18	3	3 seconds
Simulation Tests	18	6	3 seconds

Table 2.2: Implementation Parameters: Harmonic Regression Algorithm to Remove BCG Artifacts. For each test case reported, model orders and moving window lengths are listed. The model orders chosen are sometimes different from guidelines directly suggested by information criteria used in model selection (e.g, Fig. 2-4) due to reasons described in § 2.4.2.2. Regardless of these differences, we ensure that chosen model orders fulfill goodness of fit criteria. None of the test cases we studied required more than $15 - 20$ harmonics and 10 AR parameters.

Chapter 3

Subcortical Source Imaging with M/EEG and Anatomical MRI

Subcortical structures have critical roles in a variety of cognitive brain states. However, resolving electrophysiological activity in these structures remains a formidable challenge because subcortical sources, located deep in the brain, generate weak M/EEG signals. In this chapter, we introduce a novel hierarchical subspace pursuit algorithm to estimate neural currents within subcortical structures based on non-invasive M/EEG measurements. We begin by systematically comparing field patterns arising from activity in subcortical and cortical structures. These comparisons reveal that gain-insensitive information in the M/EEG field patterns offers robust distinctions between subcortical and cortical sources if sparse constraints are imposed. We then build on these observations to develop a sparse inverse solution for the subcortical source imaging problem. Specifically, we employ a subspace pursuit scheme in a hierarchical search for sparse subcortical and cortical sources underlying the measurement, and estimate fast-evolving currents in these sources to explain the data. Next, we validate the efficacy of this algorithm for recovering thalamic and brainstem contributions in challenging simulated and experimental evoked response test cases. Further, we demonstrate improved spatial resolution and specificity over existing approaches that do not build on the field pattern distinctions we have identified here. Our results establish the feasibility of resolving electrophysiologic activity within subcortical structures using non-invasive M/EEG recordings, and thus open up unique opportunities to study regional dynamics across the brain at high temporal resolution.

The M/EEG recordings and anatomical MRI data used in this work were collected in collaboration with Jyrki Ahveninen, Samantha Huang, Stephanie Rossi, Tommi Raij and Matti Hamalainen at the Massachusetts General Hospital (MGH). The subspace pursuit implementation was developed in collaboration with Gabriel Obregon-Henao at the MGH Martinos Center for Biomedical Imaging.

3.1 Introduction

Subcortical structures have critical roles in healthy and abnormal brain function. Studies of normal brain functions (sensation, memory, emotion, reward, movement, sleep and arousal) and clinically relevant brain states (dementia, depression, epilepsy, anesthesia) would benefit from the ability to access neuronal activity and regional dynamics across subcortical structures. Thus, techniques for non-invasively imaging neuronal activity in these structures with high spatiotemporal resolution are of wide interest.

While functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) resolve neuronal activity in deep brain structures with good spatial resolution, they are indirect measures as they resolve hemodynamic or metabolic correlates of neuronal activity and they have low temporal resolution; thus limiting applicability for studies of regional dynamics, circuit or network characteristics (§ 1.1). Magnetoencephalography (MEG) and electroencephalography (EEG) have the unique ability to directly measure neuronal currents at millisecond time scales [4–6], and solving the electromagnetic inverse problem to estimate the neuronal source current distribution underlying measured fields can help study regional dynamics at high temporal resolution [6, 21]. However, as subcortical structures located deep in the brain are distant from external sensors, they generate low amplitude M/EEG signals and pose formidable challenges with respect to this inverse problem [6, 106]. As such, it remains an open question as to whether electromagnetic activity in subcortical structures can be sufficiently resolved or localized with non-invasive M/EEG recordings.

Classical source localization approaches integrate (a) anatomic information from MRI, (b) tissue conductivities, and (c) Maxwell's equations to compute theoretical M/EEG field patterns for all possible source locations, and then fit these theoretical field patterns to the measured data to infer the underlying source current distribution. As the mapping between source currents and M/EEG fields is linear in the quasi-static limit, a large number of inverse solutions – employing dipole fitting [167, 168], distributed linear least squares fits with minimum norm criteria, Bayesian estimation or state-space models [6, 122, 169–172], and beam-forming approaches [121, 173] – have been applied widely and successfully to cortical sources. Some reports have assessed the applicability, resolution and accuracy of these existing techniques for deep subcortical sources. Dipole fitting approaches can correlate field patterns with expected source depths for hippocampal, thalamic and brainstem components [112–115, 117, 120, 174], but are known to become problematic in classic physiological cases having multiple active regions or highly correlated sources [21, 122]. Minimum norm estimators have been applied for subcortical source analysis, but are known to have heavy bias towards high gain superficial contributors [119, 175–177] causing poor spatial resolutions for deeper sources and requiring averages over large numbers of trials to uncover deep source activity in evoked settings [6, 106, 116, 119, 166]. Beamformer methods are typically sensitive to forward modeling errors and limited by leakage when the source space comprises weak subcortical sources alongside strong cortical sources [21, 121, 173]. Thus, existing solutions to the electromagnetic inverse problem, while

widely effective for cortical sources, have low sensitivity to deep subcortical sources – primarily as these sources contribute low amplitude (gain) signals to M/EEG measurements.

To overcome this challenge, we ask if there is gain-insensitive information in M/EEG data that can be exploited to differentiate subcortical and cortical sources. Typical M/EEG recordings comprise data from multiple sensors distributed across the head. Thus the information in these data is not only limited to absolute amplitudes of signals recorded at each individual sensor, but also includes relative amplitudes across sensors which can be thought of as gain-normalized patterns of fields across sensors. We hypothesize that constraints to distinguish subcortical vs. cortical contributions to non-invasive recordings can be found in distinctions between their respective field patterns, and that an approach built around these distinctions can localize subcortical sources.

To test this concept, we systematically compare gain-normalized field patterns arising from simulated activity in subcortical and cortical sources. These analyses reveal that field patterns originating from subcortical and cortical sources are significantly distinct when the spatially sparse nature of brain activity is taken into account. We build on these observations to propose that hierarchically reducing the space of possible sources to sparse subsets underlying a given measurement can enable subcortical source localization. The problem of identifying salient sparse subsets amongst a high-dimensional space of possibilities is addressed by projection pursuit methods. Thus we apply a subspace pursuit scheme in a hierarchical search for relevant subcortical sources. We validate efficacy on challenging simulation test-cases and auditory brainstem response recordings, and benchmark performance improvements over existing approaches. Finally, we compare the relative effectiveness of using MEG and EEG data for subcortical source localization, and conclude that combined M/EEG experimental paradigms may be most informative for these problems.

3.2 Approach: Field Pattern Comparisons

To compute gain-normalized field patterns arising from activity in subcortical and cortical regions, we (a) acquired high resolution anatomical MRIs on healthy volunteers, (b) obtained surface reconstructions and volume segmentations, (c) placed dipole sources on the cortical surface and in subcortical volumes of interest, and (d) computed electromagnetic forward solutions.

3.2.1 MRI Acquisition

Standard and high resolution T1-weighted anatomical MRIs were obtained on 5 subjects, who provided informed consent. Standard resolution MPRAGE images were obtained on a Siemens 1.5T Sonata^{TM} scanner. High resolution multi-echo MPRAGE images were obtained on a Siemens 3T TimTrio^{TM} scanner. Scanning parameters (TR=2,510 ms; 4 echoes with TEs= 1.64, 3.5, 5.36 and 7.22 ms; 176 sagittal slices, 1-mm isotropic voxels, 256 $\times$ 256 matrix; flip angle=7°) were optimized in the MGH-Martinos Center to provide maximum gray-white matter contrast for the brain reconstruction algorithms. All scans were motion-corrected and averaged to optimize SNR.

3.2.2 Source Space Construction

The FreeSurfer image analysis software was used to reconstruct and triangulate neocortical and hippocampal surfaces, as well as to segment the subcortical volumes by anatomic region [178–180]. Dipoles were positioned at vertices in the triangulated surface mesh for neocortex and hippocampus, at the gray-white matter interface, and assigned orientations fixed normal to the neocortical and hippocampal surfaces respectively. The neocortical and hippocampal dipole placements are performed consistently as both structures have a similar layered cytoarchitecture, and as the hippocampal inner structure is not accessible with anatomical MRI [119]. Dipoles were also placed in subcortical volumes (thalamus, caudate, putamen, amygdala, medial and lateral geniculate nuclei, brainstem) extracted from the FreeSurfer segmentations, as well as in specific smaller nuclei (medial and lateral geniculate) defined based on standard anatomic landmarks and FreeSurfer segmentations, at 1 mm voxel spacing, and oriented randomly as there is no known preferred orientation for dipoles in these regions. This source space comprises $\sim 10^6$ dipoles across the brain. But with only $\sim 10^2$ sensors and highly correlated activity of neighboring dipoles, it is difficult to independently resolve dipole activity to 1 mm. Therefore we grouped dipoles into "surface patches" for neocortex and hippocampus, and "volume subdivisions" for subcortical volumes.

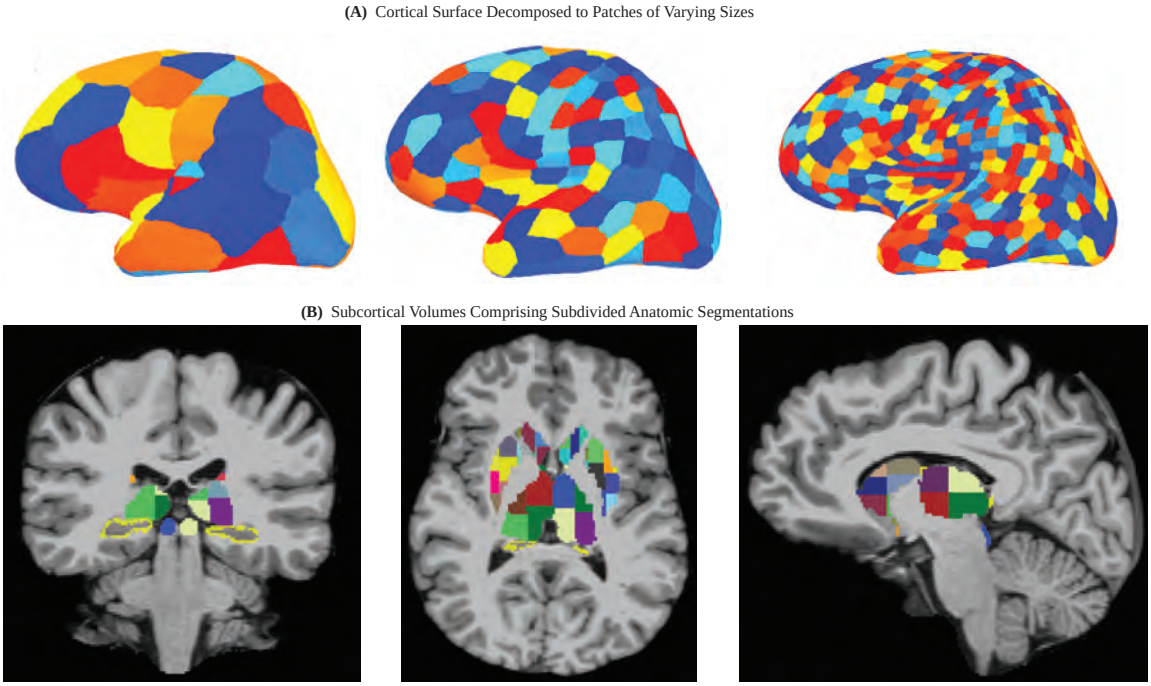

Figure 3-1: Cortical and Subcortical Source Spaces. (A) Cortical source space comprising patches on the gray-white matter surface interface (figure adapted from [86]). The set of coarse patches with average area $2500\,\mathrm{mm}^2$ (on left) is denoted $\mathbb{C}^{(1)}$, the set of finer patches with average area $650\,\mathrm{mm}^3$ (middle) is denoted $\mathbb{C}^{(2)}$, and the set of even finer patches with average area $175\,\mathrm{mm}^2$ (right) is denoted $\mathbb{C}^{(3)}$. *(B)* Subcortical source space $\mathbb{S}$ comprising volume subdivisions ranging $150 - 800\,\mathrm{mm}^3$ across caudate, putamen, hippocampus, amygdala, thalamus, lateral and medial geniculate nuclei, and brainstem.

Cortical and hippocampal dipoles were grouped into patches within the MNE software package [82, 181]. The cortical and hippocampal surfaces were each approximated with the topology of recursively subdivided icosahedra. For a given icosahedron subdivision, detailed surface geometry information was used to derive "surface patches". Subcortical dipoles were grouped into "volume subdivisions" obtained by subdividing the FreeSurfer-based anatomic subcortical segmentations. The volume subdivisions, agnostic to specific anatomy within the segmentations, were sized to homogenize current strengths across regions as follows. First, volume current densities ρ were specified in a structure-specific manner [119]. Second, using neocortical surface current density σ and average cortical patch area $\bar{A}$, the average current strength for cortical patches was computed as $\sigma\bar{A}$. Third, total volumes V for anatomically distinct subcortical regions were computed. The number of subdivisions for a given subcortical volume segmentation was determined as $n = \rho V/\sigma\bar{A}$. Then, the subcortical segmentation was subdivided into $\sqrt[3]{n}$ equal volumes along each of the 3 axes. The typical volume of a subcortical subdivision ranges $150 - 800\,\mathrm{mm}^3$, depending on structure (§ 3.8.1). This process of homogenizing current strengths implicitly assigns higher resolution (finer subdivisions) to regions with high current densities and lower resolution (larger subdivisions) to regions with low current densities. The resulting cortical and subcortical source spaces are shown in Fig. 3-1.

3.2.3 Electromagnetic Forward Solutions

Activity in dipoles within a given region, denoted by source currents $\mathbf{x}$, generates the resulting field patterns $\mathbf{y}$ at the sensors via Maxwell's equations encoded in the electromagnetic forward solution $\mathbf{G}$ for that region:

$$\mathbf{y}_{N\times 1} = \mathbf{G}_{N\times M}\mathbf{x}_{M\times 1}, \tag{3.1}$$

where N is the number of sensors and M is the number of dipole sources in the region of interest. The mapping specified in $\mathbf{G}$ is linear as it arsies from a quasi-static approximation to Maxwell's equations. MEG forward solutions employed the sensor configuration of a Neuromag Vectorview™ MEG system (Elekta Neuromag, Helsinki, Finland) comprising 306 sensors across 102 sensor locations (2 planar gradiometers and 1 magnetometer per location). EEG forward solutions employed a standard 70-lead EEG electrode configuration employed at MGH and in the MNE software Sample dataset. The positions of MEG and EEG sensors and the fiduciary points were coregistered with the subject's MRI data using standard procedures (§ 3.4.1). For each regional subdivision in the source space, MEG and EEG forward solutions were generated within the MNE software package [82, 181], using a three-compartment boundary element model derived from the MRI data. Dipoles within 5 mm of the inner skull bounding surface were excluded to avoid numerical errors. As the current densities across regions in this combined subcortical-cortical source space are non-homogenous [106], the $\mathbf{G}$ for each region was multiplied by regional current strength denoted as c. For each subcortical region, current strength was derived as $c = \rho V$ where ρ and V are the regional current density (from [106]) and volume respectively. For each neocortical and hippocampal patch,

current strength was derived as $c = \sigma A$ where σ and A are the surface current density (from [106]) and area respectively.

For any given subdivision of the source space, the columns of $\mathbf{G}$ represent fields arising from small collections of neighboring dipoles and thus are highly correlated/degenerate. Thus, information in the forward solutions for individual regions can be condensed into a low-dimensional approximation i.e., into a low-rank basis derived using a reduced singular value decomposition [86, 182]:

$$\mathbf{G} = \mathbf{U}_{N\times B}\mathbf{S}_{B\times B}\mathbf{W}_{B\times M}, \tag{3.2}$$

$$\hat{\mathbf{G}} = \mathbf{U}_{N\times P}\mathbf{S}_{P\times P}, \tag{3.3}$$

where $P < B$ is the number of eigenmodes retained in the low-rank approximation. For each subdivision, P is chosen such that the reduced order $\hat{\mathbf{G}}$ represents at least 95% of the total spectral energy in $\mathbf{G}$:

$$\frac{\sum_{b=1}^{P}\mathbf{S}_{bb}}{\sum_{b=1}^{B}\mathbf{S}_{bb}} \geq 0.95.$$

Typically $P = 2-6$ modes are sufficient to capture 95% of the spectral energy in $\mathbf{G}$. This approach (a) enables reduction of the dimensionality of $\mathbf{G}$, (b) minimizes representation error on average [182], and (c) implicitly defines reduced subspaces for the forward solution from each region, for use in field comparisons and inverse solutions.

3.2.4 Gain Normalized Field Patterns

The field patterns arising from activity in a region of interest can be gain-normalized by considering the eigenmodes of $\mathbf{G}$ corresponding to that region of interest, and activating each mode with currents proportional to the fraction of total spectral energy contained within the mode:

$$\mathbf{y}_{\text{norm}} = \mathbf{G}_{\text{norm}}\mathbf{x}_{\text{norm}} = \mathbf{U}_{N\times P}\frac{\mathbf{S}_{P\times P}}{\sum_{p=1}^{P}\mathbf{S}_{pp}} = \hat{\mathbf{G}}\frac{1}{\sum_{p=1}^{P}\mathbf{S}_{pp}}, \tag{3.4}$$

where $\hat{\mathbf{G}}$, $\mathbf{U}$ and $\mathbf{S}$ are as in Eqn. 3.2-Eqn. 3.3, and $\mathbf{y}_{\text{norm}}$ represents the normalized MEG field patterns generated by unit activity within a region of interest, assuming no noise, at one snapshot in time for the simplest analysis.

3.2.5 Principal Angle Computations

Every $\mathbf{G}_{\text{norm}}$ or $\hat{\mathbf{G}}$ embodies the subspace of all possible field patterns $\mathbf{y}_{\text{norm}}$ that can arise from some source current distribution ($\mathbf{x}_{\text{norm}}$) within a region. Thus, distinctions between the possible field patterns from different regions can be quantified using metrics to quantify distinctions between their respective subspaces. Principal angles are established metrics for subspace comparisons.

Formally, principal angles between subspaces of possible field patterns from regions i and j, defined by the reduced order electromagnetic forward models $\hat{\mathbf{G}}_i$ and $\hat{\mathbf{G}}_j$ respectively, are denoted as $\theta_{i,j}$. The angles can be computed using a formulation that uses the singular value decomposition (described in [183]):

$$\hat{\mathbf{G}}_i'\hat{\mathbf{G}}_j = \mathbf{U}\mathbf{p}_{i,j}\mathbf{\Sigma}_{i,j}\mathbf{W}\mathbf{p}_{i,j}, \tag{3.5}$$

$$\mathbf{\Sigma}_{i,j} = \mathrm{diag}\left(\mathrm{s}_1, \mathrm{s}_2, \ldots \mathrm{s}_\mathrm{n}\right), \tag{3.6}$$

$$= \mathrm{diag}\left(\cos\theta_1, \cos\theta_2, \ldots \cos\theta_\mathrm{n}\right), \tag{3.7}$$

$$\Theta_{i,j} = \left[\theta_1, \theta_2, \ldots \theta_n\right]. \tag{3.8}$$

Intuitively, low singular values in $\Sigma_{i,j}$ indicate uncorrelated eigenmodes and correspond to higher angles in $\Theta_{i,j}$, whereas high singular values indicate correlated eigenmodes and lower angles. Thus, low angles ($\sim 0°$) suggest that the subspaces defining field patterns from regions i and j have nearly parallel components with one subspace a near subset of the other, while high angles ($\sim 90°$) suggest that the subspaces are nearly orthogonal or have a null intersection.

Analyses were performed using the implementation described in [184]. As each pair corresponds to n angles where n is the minimum of the ranks of $\hat{\mathbf{G}}_i$ and $\hat{\mathbf{G}}_j$, the means of $\Theta_{i,j}$, denoted as $\bar{\Theta}_{i,j}$, are reported. For notational convenience, we will denote the low-rank approximations $\hat{\mathbf{G}}$ as $\mathbf{G}$ in the following sections.

3.2.6 Distinctions in Field Patterns from Sparse Sources

Fig. 3-2*A* shows an example of a subcortical volume source space of interest $\mathbb{S}$, in ventral posterior lateral thalamus or VPL, with gain-normalized forward solution denoted by $\mathbf{G}_\mathbb{S}$. Unit activity in the dipoles of $\mathbb{S}$ generates, via Eqn. 3.4, the gain-normalized field pattern across sensors $\mathbf{y}_\mathbb{S}$, shown in Fig. 3-2*B*. We first asked if activity in dipoles of the cortical source space $\mathbb{C}^{(3)}$, denoted for convenience as $\mathbb{C}$ (shown in Fig. 3-2*C*) can generate the subcortical field pattern $\mathbf{y}_\mathbb{S}$. To assess this, we fit the gain-normalized cortical forward solution $\mathbf{G}_\mathbb{C}$ to the subcortical field $\mathbf{y}_\mathbb{S}$ using the least squares criterion. The minimum norm fit $\mathbf{y}_\mathbb{C} = \mathbf{G}_\mathbb{C}\mathbf{x}_\mathbb{C} = \mathbf{G}_\mathbb{C}\left(\mathbf{G}_\mathbb{C}'\mathbf{G}_\mathbb{C}\right)^{-1}\mathbf{G}_\mathbb{C}'\mathbf{y}_\mathbb{S}$ is shown in Fig. 3-2*D*. It appears that the full cortical source space can explain the subcortical field pattern $\mathbf{y}_\mathbb{S}$.

But as brain activity often exhibits spatiotemporal sparsity, it is unlikely that the full cortical source space $\mathbb{C}$ is active at any given time. Thus we asked if subcortical field pattern $\mathbf{y}_\mathbb{S}$ is distinguishable from cortical field patterns when a physiologically relevant sparse cortical region is active. Fig. 3-2*E* shows a source space $\mathbb{C}_{\mathrm{sp}}$ in a patch of somatosensory cortex, which is often co-active with the VPL thalamus sources $\mathbb{S}$ in somatic sensation. We asked if the gain-normalized forward solution for $\mathbb{C}_{\mathrm{sp}}$, i.e. $\mathbf{G}_{\mathbb{C},\mathrm{sp}}$ can explain the subcortical field $\mathbf{y}_\mathbb{S}$. The least squares fit this time, $\mathbf{y}_{\mathbb{C},\mathrm{sp}} = \mathbf{G}_{\mathbb{C},\mathrm{sp}}\mathbf{x}_{\mathbb{C},\mathrm{sp}} = \mathbf{G}_{\mathbb{C},\mathrm{sp}}\left(\mathbf{G}_{\mathbb{C},\mathrm{sp}}'\mathbf{G}_{\mathbb{C},\mathrm{sp}}\right)^{-1}\mathbf{G}_{\mathbb{C},\mathrm{sp}}'\mathbf{y}_\mathbb{S}$, is in Fig. 3-2*F*. The fit is poor - suggesting that sparse cortical regions cannot generate the same field pattern as some subcortical region.

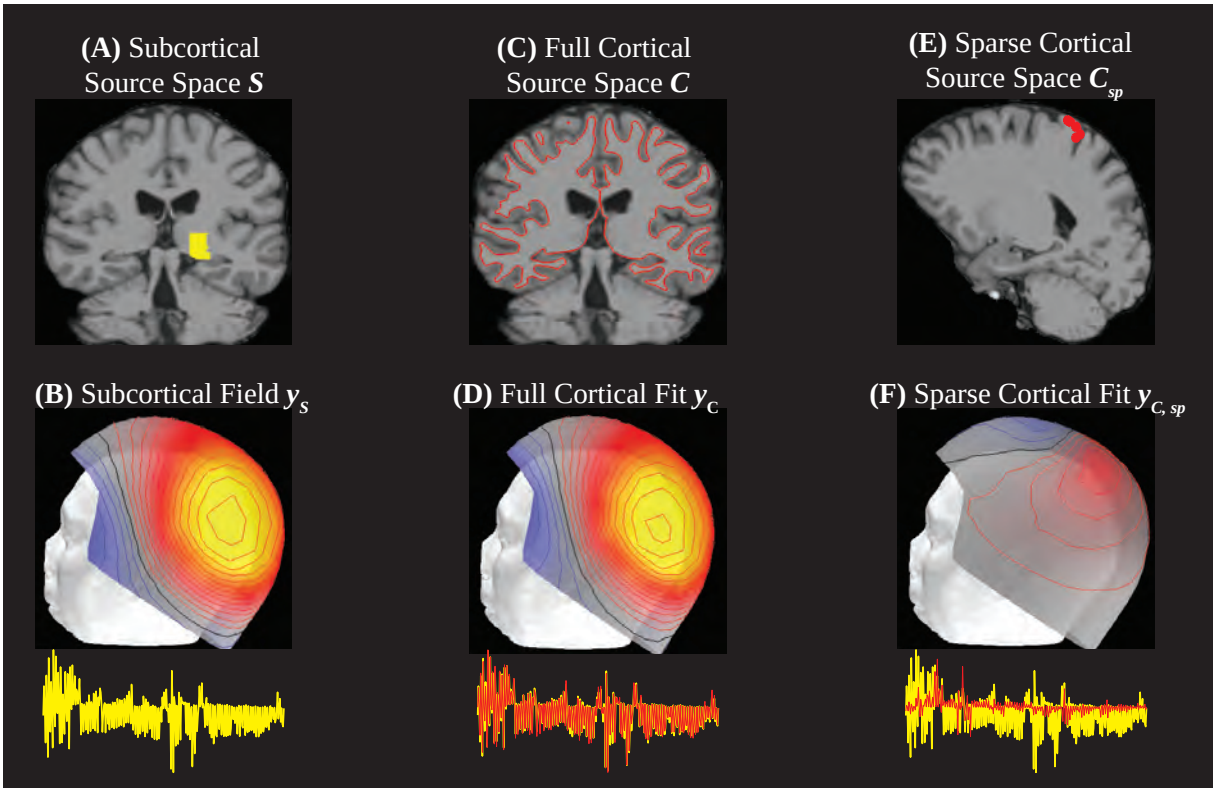

Figure 3-2: Sparse Cortical Regions Cannot Explain Field Patterns Generated by Subcortical Regions. The full cortical source space can capture MEG field patterns generated in a subcortical region but sparse cortical source spaces cannot. *(A)* Subcortical source space $\mathbb{S}$ comprising discrete dipoles in the volume of ventral posterior lateral thalamus. The corresponding forward solution is $\mathbf{G}_{\mathbb{S}}$. *(B)* The field pattern $\mathbf{y}_{\mathbb{S}} = \mathbf{G}_{\mathbb{S}}\mathbf{x}_{\mathbb{S}}$ where $\mathbf{x}_{\mathbb{S}}$ is gain-normalized unit activity. *(C)* Full cortical surface source space $\mathbb{C}$ with forward solution $\mathbf{G}_{\mathbb{C}}$. *(D)* The field pattern $\mathbf{y}_{\mathbb{C}}$ obtained upon fitting $\mathbf{G}_{\mathbb{C}}$ to $\mathbf{y}_{\mathbb{S}}$. *(B vs. D)* The full cortical source space can explain the subcortical field pattern. *(E)* A sparse cortical surface patch $\mathbb{C}_{\text{sp}}$ with forward solution $\mathbf{G}_{\mathbb{C},\text{sp}}$. *(F)* The field pattern $\mathbf{y}_{\mathbb{C},\text{sp}}$ obtained upon fitting $\mathbf{G}_{\mathbb{C},\text{sp}}$ to $\mathbf{y}_{\mathbb{S}}$. *(B vs. F)* The sparse cortical source space cannot explain the subcortical field pattern. As we are plotting fits to subcortical amplitude, and the fit is poor, fields in Panel *F* have lower amplitude than fields in Panel *B*. *(B,D,F)* All field patterns were generated within the MNE Software Package with a contour step of 5×10^{-4}, and are shown both topographically across the MEG helmet and as raw measurement series across sensors. For the topographic maps - yellow denotes positive maximum and blue denotes negative minimum. For the measurement series plots - yellow series is from $\mathbb{S}$, and red series denotes the cortical fits. Field patterns are gain-normalized and thus unitless, so as to highlight the pattern of relative amplitudes across sensors (instead of absolute amplitudes at any given sensor).

We note that this result has purely anatomic and physical origins as the columns of $\mathbf{G}$ are solely a function of source and sensor locations, head geometry and tissue conductivity properties. Further, as this analysis corresponds to gain-normalized field patterns $\mathbf{y}_{\mathbb{S}}$, it is insensitive to differences in amplitudes of fields generated by subcortical and cortical sources. Thus, the poor fit in Fig. 3-2*F* must be because the field arising from activity in a sparse brain region spans only a limited area on the MEG helmet and thus cannot capture the entire space of possible field patterns generated by activity in another sparse region. In other words, it must arise because the subspace spanned by $\mathbf{G}_{\mathbb{S}}$ does not have significant overlap with the subspace spanned by $\mathbf{G}_{\mathbb{C},\text{sp}}$.

3.2.7 Comparisons between Subspaces of Field Patterns

To quantify the degree of overlap between subcortical and cortical field patterns, we computed principal angles $\bar{\Theta}_{i,j}$ for all pairs (i, j) of cortical and subcortical regions.

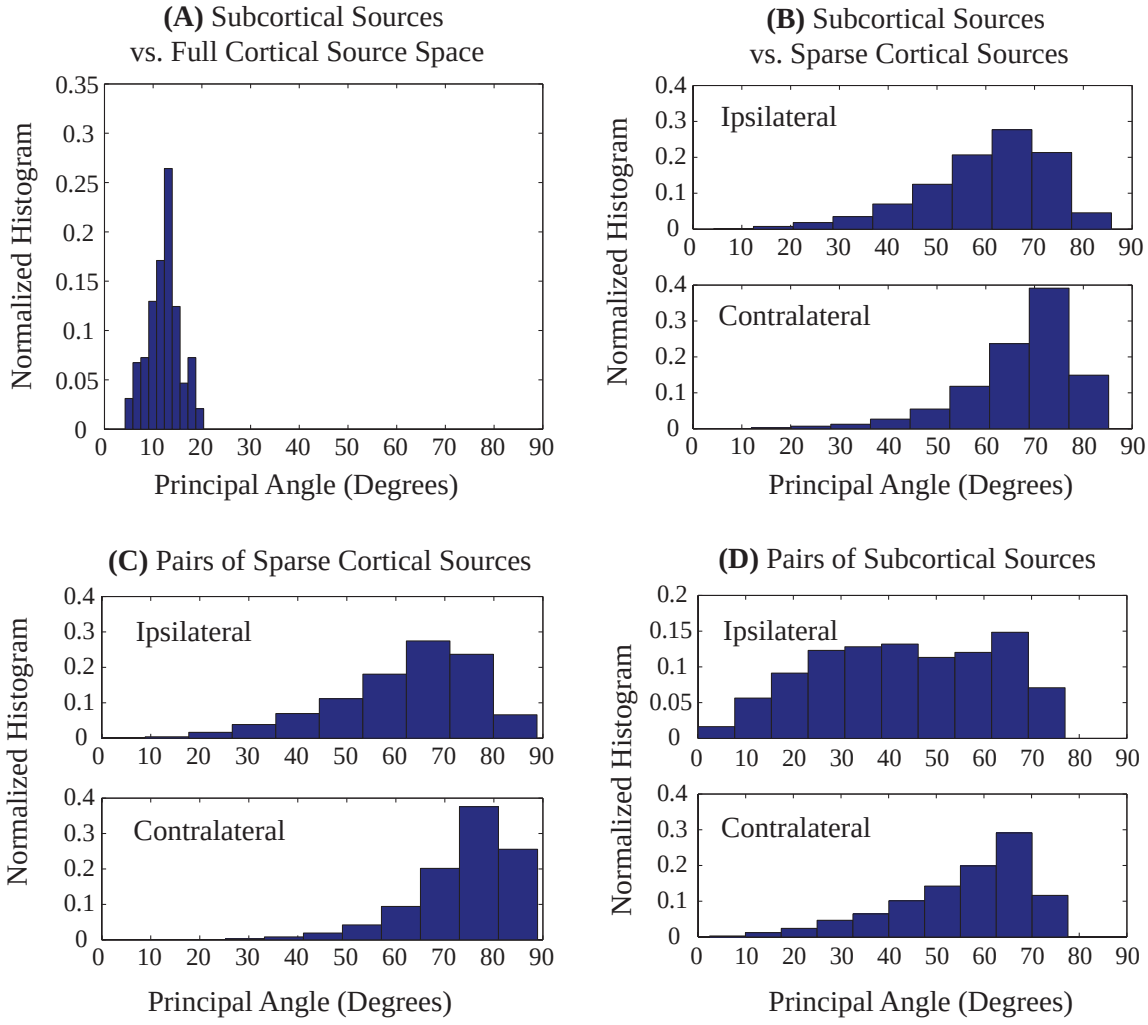

Figure 3-3: Sparse Brain Regions Generate MEG Field Patterns in Distinguishable Subspaces. Normalized histograms of principal angles between pairs of brain regions for MEG field patterns. Angles quantify distinctions between subspaces defining field patterns from the two regions. *(A vs. B)* Full cortical source space can generate most any field pattern that a subcortical source space can, while sparse cortical sources cannot. *(B vs. C)* Sparse subcortical and cortical source spaces generate field patterns that are just as distinct as those generated by pairs of sparse cortical source spaces. *(C vs. D)* Pairs of subcortical regions also generate distinct field patterns, like pairs of cortical regions, but exhibit a bit more overlap due to their greater distance from sensors. *(B-D)* A large proportion of the principal angles are high and close to orthogonal when sparse regions are compared. All field pattern comparison results here are shown for $\mathbb{S}$ and $\mathbb{C}^{(3)}$ but hold for coarser cortical source spaces too (Fig. 3-11).

Fig. 3-3*A* shows the principal angles between field patterns arising from activity in the subcortical regional subdivisions and the full cortical source space respectively. The angles are low suggesting that the full cortical source space can explain most field patterns arising from a subcortical subdivision (as also seen in Fig. 3-2*A*-*D*). This is because the full cortical case has several eigenmodes spanning most possible field patterns – including those generated by activity within a subcortical volume subdivision. To mimic the test cases in Fig. 3-2, we compare angles between field patterns from sparse subcortical and cortical regions (Fig. 3-3*B*). The angles in Fig. 3-3*B* are significantly higher than those in Fig. 3-3*A,* suggesting near orthogonality between field patterns generated by sparse pairs of regions (as also seen in Fig. 3-2*E*-*F*). To interpret the results in Fig. 3-3*B* more carefully, we compare them with a "control case", i.e., the field patterns generated by pairs of sparse cortical regions. Fig. 3-3*C* shows that pairs of cortical regions, too, give rise to near orthogonal field patterns. Further, the angle trends in Fig. 3-3*B*-*C* are highly similar - suggesting that when viewed within the information space of gain-normalized field patterns, the problem of distinguishing subcortical and cortical sources is just as solvable as the problem of distinguishing cortical sources from each other.

Next, we quantify distinctions between field patterns arising from pairs of subcortical regions. Fig. 3-3*D* has similar trends as in Fig. 3-3*C* , albeit with a bit more overlap between field patterns for pairs of subcortical regions than for pairs of cortical regions. Difference in trends between Fig. 3-3*C*-*D* can be understood using an analogy with the simple case of two dipole sources and a few sensors. When the dipole sources are close to each other, but far away from the sensors (for e.g., pairs of subcortical regions), they appear quite similar to the sensors. However, when the dipole sources are close to each other, and just as close to the sensors (for e.g., pairs of cortical regions), then they will appear more distinct to the sensors. Regardless, there are still several pairs of subcortical regions in Fig. 3-3*D* which have highly distinct field patterns. Finally, we note that the field patterns arising from pairs of contralateral regions are typically more orthogonal than those arising from pairs of ipsilateral regions (Fig. 3-3*B*-*D*). This suggests that it will be easier to resolve subcortical activity in tasks where contralateral subcortical regions are involved. Yet, by no means is this a requirement as in Fig. 3-3*B* and *D*, even the ipsilateral regions contain highly orthogonal field patterns.

In summary, in every case where sparse regions are compared, the field patterns are more distinguishable than the naive case of Fig. 3-3*A*. Overall, these results illustrate, with minimal assumptions, that the gain-normalized field arising from activity in sparse subcortical and cortical regions cannot map identically to all sensors. Rather, these comparisons reveal robust distinctions in gain-insensitive field patterns generated by sparse subcortical and cortical regions across the brain. Coincidentally, these distinctions lend themselves well to the natural spatiotemporal dynamics of brain activity in physiologic and pathologic conditions. Thus, we sought to exploit these distinctions in an inverse framework to assess whether subcortical sources can be localized.

3.3 Inverse Algorithm

3.3.1 Electromagnetic Inverse Problem

The electromagnetic inverse problem is to estimate the source currents $\mathbf{X}$ underlying M/EEG measurements $\mathbf{Y}$, given the electromagnetic forward solution $\mathbf{G}$ linking the two:

$$\mathbf{Y}_{N\times T} \;=\; \mathbf{G}_{N\times M}\mathbf{X}_{M\times T} + \mathbf{V}_{N\times T}, \tag{3.9}$$

where the observation noise $\mathbf{V}$ is assumed to be Gaussian, temporally uncorrelated, and specified completely by spatial covariance $\mathbf{Q}_{N\times N}$ independent of $\mathbf{X}$, N is the number of sensors, M is now the number of dipole sources across the brain, and T is the number of time points in the measurement.

3.3.2 Hierarchical Approach for Subcortical Source Estimation

The classic solution for the linear inverse problem of estimating $\mathbf{X}$ given measurements $\mathbf{Y}$, forward solution $\mathbf{G}$, noise statistics $\mathbf{Q}$ and a prior covariance for the source currents $\mathbf{R}$ is given by the minimum norm estimate (MNE):

$$\hat{\mathbf{X}}^{MNE}(\mathbf{Y},\mathbf{G},\mathbf{Q},\mathbf{R}) \;=\; \mathbf{R}\mathbf{G}'\Big(\mathbf{G}\mathbf{R}\mathbf{G}'+\mathbf{Q}\Big)^{-1}\mathbf{Y}, \tag{3.10}$$

which is essentially the projection of measurement $\mathbf{Y}$ onto the range of $\mathbf{G}$ while accounting for noise characteristics and prior information. Minimum norm estimation is the most widely used source imaging technique, and reliably localizes cortical sources underlying the M/EEG measurements.

However, our field comparison analysis suggests that a simple application of MNE on the full brain, with $\mathbf{G}$ specifying the composite forward solution for all subcortical and cortical sources, may not reveal subcortical sources amidst a full cortical source space, as the latter can almost always explain activity in the former. Thus, the naive MNE estimate is likely to localize to the cortical source space even if the activity arises from a subcortical source. Instead, as distinctions in subcortical and cortical field patterns arise when sparse source spaces are considered, localizing subcortical contributions requires narrowing the space of possible cortical sources to relevant sparse subsets. This will implicitly decorrelate the problem of uncovering subcortical sources amidst cortical sources. Then, the estimation can be repeated with a combination of this reduced cortical space and the full subcortical space to identify subcortical contributions.

This approach of hierarchically reducing source spaces to subsets most relevant to the measurement is illustrated in Fig. 3-4. It is possible to search the space of all possible sources (Fig. 3-4*A*), find a sparse cortical solution (Fig. 3-4*B*), perform a finer search in the neighborhood of this sparse cortical solution (Fig. 3-4*C*), obtain a refined sparse cortical solution (Fig. 3-4*D*) which is distinguishable from subcortical sources in general, and then search for subcortical sources amidst

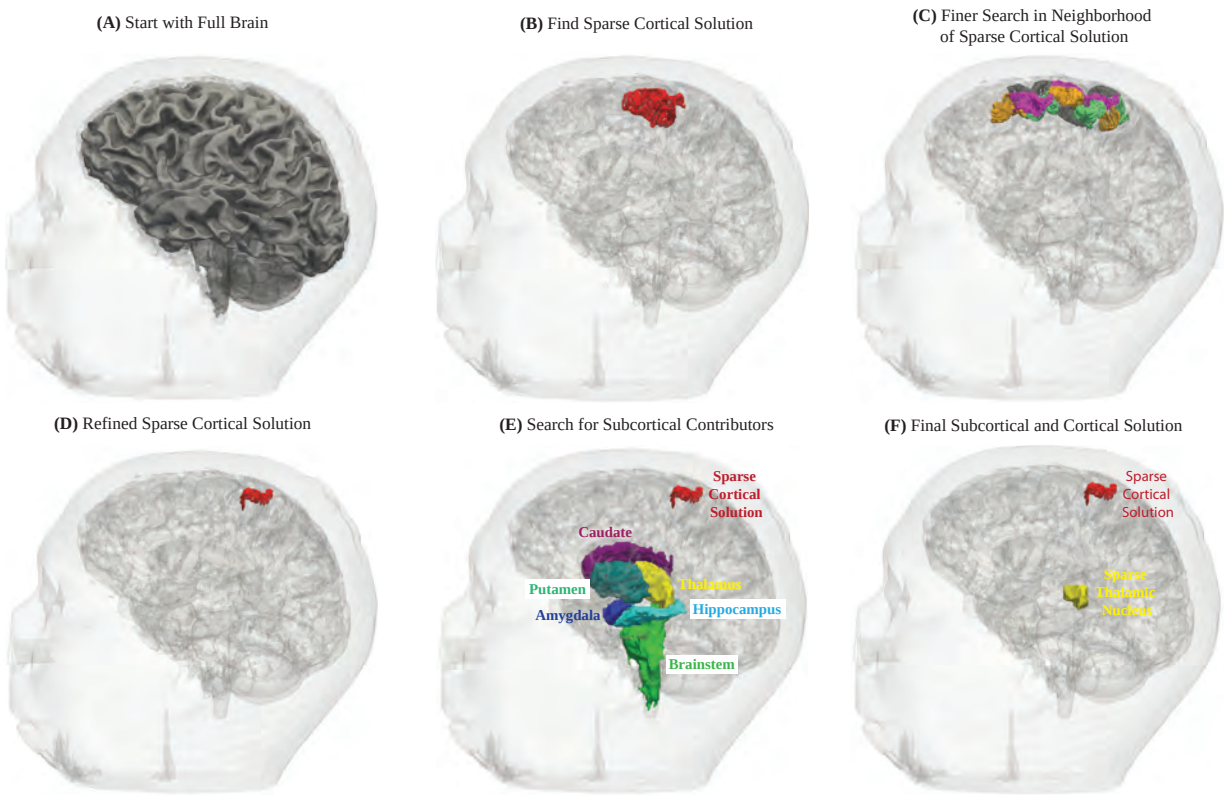

Figure 3-4: Hierarchical Reduction of Source Spaces to Relevant Sparse Subsets with Distinguishable Field Patterns. (*A*) The space of all possible sources across the brain. *(B)* A sparse cortical region most relevant to the measurement under study. *(C)* Refined sparse cortical regions within and around the region in Panel *B*. Different colors are used for neighboring patches to highlight patch boundaries. *(D)* Refined sparse cortical region most relevant to the measurement. *(E)* Joint source space comprising the sparse cortical region and subcortical volumes. The sparse subset of the cortical source space allows subcortical areas to show through in the inverse solution. It is possible to search across these sparse and therefore distinguishable sets of sources to localize appropriate subcortical and cortical sources. *(F)* Final sparse cortical and subcortical regions most relevant to the measurement under study.

the reduced cortical source space (Fig. 3-4*E*) to identify relevant subcortical and cortical sources (Fig. 3-4*F*).

3.3.3 A Suitable Class of Algorithms

An ideal algorithm suited to such an approach would have the ability to (a) refine the MNE to obtain accurate sparse cortical estimates and (b) search for subcortical sources amidst a reduced cortical space. Projection pursuit methods [185, 186], which identify sparse projections explaining multivariate data, meet both objectives well. First, pursuit methods can find sparse cortical regions most relevant to the measured fields, amidst the high-dimensional dictionary of all cortical regions and their corresponding field patterns. This is because they deal uniquely well with the curse of dimensionality. Second, upon identifying the sparse cortical contributors, the pursuit search can be repeated in the reduced source space to search for weak but distinct subcortical contributors. This is

because pursuit methods are robust to noise and outliers, and can recover small amplitude features in large dictionaries ("needle in haystack") with near-orthogonal columns [186]. These methods can be implemented with subspace pursuit algorithms [187, 188], which find a sparse projection matching a given dictionary (e.g., matching in the least squares or MNE sense), remove the component along that projection and iterate to find new projections, till all matching projections have been found. Subspace pursuit algorithms have recently been applied for cortical source localization [86], and thus provide a good starting point for our purpose.

Intuitively, subspace pursuit offers a principled way to iteratively pick out the "relevant" projections or rows of $\hat{\mathbf{X}}^{MNE}$ and refine the MNE estimate until all "relevant" projections have been found. Formally, for a distributed source space comprising K regions across the brain, columns of $\mathbf{G}$ and rows of $\mathbf{X}$ can be grouped by region and Eqn. 3.9 can be rewritten as:

$$\mathbf{Y}_{N\times T} = [\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_K] \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_K \end{bmatrix} + \mathbf{V}_{N\times T}, \tag{3.11}$$

where the set of regions (patches or subdivisions) is $[1, 2, \ldots, K]$, $\mathbf{G}_k$ is the electromagnetic forward solution for the k^{th} region, and $\mathbf{x}_k$ denotes the source current of dipoles in the k^{th} region across time. Then the subspace pursuit (SP) estimate is written as:

$$\left[\mathcal{H},\ \hat{\mathbf{X}}_{\mathcal{H}}\right] = \text{SP}\left(\mathbf{Y},\ \mathbf{G},\ \mathbf{Q},\ \mathbf{R}\right), \tag{3.12}$$

where $\mathcal{H}$ is a set of indices denoting the subset of regions deemed to be most relevant to the measurement. Implicitly $\mathcal{H}$ gives a subset of regions whose normalized field patterns match those in the measurement after compensating for noise, and $\hat{\mathbf{X}}_{\mathcal{H}}$ specifies source current estimates that can generate these different field patterns to quantitatively explain measurement amplitudes.

3.3.4 Hierarchical Subspace Pursuit Algorithm

We compute minimum norm estimates per [122, 189] and adapt the subspace pursuit implementation in [86] across the hierarchy of source spaces for subcortical source localization. We start with a coarse cortical source space $\mathbb{C}^{(1)}$, with forward model denoted as $\mathbf{G}_{\mathbb{C}^{(1)}}$ and perform subspace pursuit to pick L regions that can best explain measured fields:

$$\left[\mathcal{H}_1,\ \hat{\mathbf{X}}_{\mathcal{H}_1}\right] = \text{SP}\left(\mathbf{Y},\ \mathbf{G}_{\mathbb{C}^{(1)}},\ \mathbf{Q},\ \mathbf{R},\ L\right), \tag{3.13}$$

where $\mathcal{H}_1$ denotes the chosen set of L regions in $\mathbb{C}^{(1)}$ (Fig. 3-4B) and $\hat{\mathbf{X}}_{\mathcal{H}_1}$ denotes the source current estimates in these chosen regions. Next, we repeat the estimation in a finer cortical source space $\mathbb{C}^{(2)}$ - with patches overlapping $\mathcal{H}_2$ and their nearest neighbors (Fig. 3-4C). The forward model for these patches is denoted as $\mathbf{G}_{\mathbb{C}^{(2)}}$ and we perform subspace pursuit to pick L regions that can best

explain measured fields:

$$\left[\mathcal{H}_2,\ \hat{\mathbf{X}}_{\mathcal{H}_2}\right] = \mathrm{SP}\left(\mathbf{Y},\ \mathbf{G}_{\mathbb{C}^{(2)}},\ \mathbf{Q}\,,\mathbf{R},L\right), \tag{3.14}$$

where $\mathcal{H}_2$ denotes the chosen set of L regions in $\mathbb{C}^{(2)}$ (Fig. 3-4*D*) and $\hat{\mathbf{X}}_{\mathcal{H}_2}$ denotes the source current estimates in these chosen regions. This systematic reduction of possible cortical sources can be repeated to the level of fineness relevant to cortical activity in the task under study (typically $\mathbb{C}^{(3)}$). This hierarchical reduction process results in a highly spatially resolved subset of relevant sparse cortical sources, denoted as $\mathbb{C}_{\mathrm{sp}}$ (e.g. as in Fig. 3-4*D*). Then, in the final stage search, the subcortical source space $\mathbb{S}$, containing volume subdivisions of caudate, putamen, hippocampus, amygdala, thalamus, lateral and medial geniculate nuclei, brainstem, is considered alongside the sparse cortical source space $\mathbb{C}_{\mathrm{sp}}$ (Fig. 3-4*E*). We denote the composite or joint source space as $\mathbb{J} = [\mathbb{C}_{\mathrm{sp}}, \mathbb{S}]$. The forward model for $\mathbb{J}$, denoted as $\mathbf{G}_{\mathbb{J}} = [\mathbf{G}_{\mathbb{C},\mathrm{sp}},\ \mathbf{G}_{\mathbb{S}}]$, implicitly contains nearly decorrelated columns, and we perform subspace pursuit to pick L regions that can best explain measured fields:

$$\left[\mathcal{H}_{\mathbb{J}},\ \hat{\mathbf{X}}_{\mathcal{H}_{\mathbb{J}}}\right] = \mathrm{SP}\left(\mathbf{Y},\ \mathbf{G}_{\mathbb{J}},\mathbf{Q},\ \mathbf{R},\ L\right), \tag{3.15}$$

where $\mathcal{H}_{\mathbb{J}}$ specifies the set of chosen regions whose dipole activity best explains the data (e.g. as in Fig. 3-4*F*), while $\hat{\mathbf{X}}_{\mathcal{H}_{\mathbb{J}}}$ gives the source current estimates in these subdivisions. We denote these dipole currents as $\hat{\mathbf{X}}$ for notational convenience.

3.4 Data Acquisition and Analysis

3.4.1 M/EEG Acquisition

MEG and EEG data were acquired simultaneously within an electromagnetically shielded room at the Athinoula A. Martinos Center for Biomedical Imaging (Charlestown, MA USA), using a 306-channel dc-SQUID Neuromag Vectorview system (Elekta-Neuromag, Helsinki, Finland) and a 70-lead EEG cap. Electrooculograms (EOG) were also recorded to monitor blink artifacts. The position of the subject's head was registered with respect to the MEG SQUID sensors with HPI coils, and to the EEG sensor locations with the FastTrak 3D digitizer (Polhemus, Inc., Colchester, VT, USA) respectively. All M/EEG acquisitions were accompanied with MRI acquisitions (detailed in § 3.2), and head and sensor coordinates were aligned to MRI data with digitized anatomic landmarks.

3.4.2 Data Analysis Overview

The procedure to analyze a given dataset with the hierarchical subspace pursuit algorithm is very similar to the standard procedure for source localization with MNE. The steps in order are (a) construct source spaces, (b) compute forward solutions $\mathbf{G}$, (c) preprocess measurements $\mathbf{Y}$ and estimate noise covariances $\mathbf{Q}$, (d) specify source covariance $\mathbf{R}$ and compute the MNE estimates, (e)

refine these estimates with subspace pursuit, and (f) perform pursuit across hierarchies to analyze the data. Steps (a-b) were detailed in § 3.2. Step (c) will be described for each individual test case. We now describe the algorithmic steps for (d)-(e), and detail the data analysis choices for (f).

3.4.3 Minimum Norm Estimate

Standard procedures specified in [122, 189] were used to compute the minimum l_2-norm estimate (MNE) for the electromagnetic inverse problem. The MNE solution for source currents $\mathbf{X}$ given full forward solution $\mathbf{G}$, M/EEG measurements $\mathbf{Y}$, noise statistics $\mathbf{Q}$ and prior covariance $\mathbf{R}$ on source currents is given by:

$$\hat{\mathbf{X}}^{MNE}(\mathbf{Y},\mathbf{G},\mathbf{Q},\mathbf{R}) = \mathbf{R}\mathbf{G}'\left(\mathbf{G}\mathbf{R}\mathbf{G}'+\mathbf{Q}\right)^{-1}\mathbf{Y}. \tag{3.16}$$

The observation noise covariance $\mathbf{Q}$ is estimated from the data (or given as a known matrix in the case of simulations). With the eigen-decomposition of $\mathbf{Q} = \mathbf{U_Q}\mathbf{\Lambda_Q^2}\mathbf{U_Q'}$, it is possible to write $\mathbf{Q}^{-1/2} = \Lambda_\mathbf{Q}^{-1}\mathbf{U_Q'}$. Premultiplying the measurement model by $\mathbf{Q}^{-1/2}$ gives the whitened measurement $\tilde{\mathbf{Y}} = \mathbf{Q}^{-1/2}\mathbf{Y}$ and whitened forward model $\tilde{\mathbf{G}} = \mathbf{Q}^{-1/2}\mathbf{G}$. The dimensionality of the problem was reduced by using the low-rank approximations of the whitened forward models $\tilde{\mathbf{G}}_{\mathrm{low-rank}}$ in lieu of the full $\tilde{\mathbf{G}}$ (per Eqn. 3.3). For notational convenience, we will denote the low-rank approximations $\tilde{\mathbf{G}}_{\mathrm{low-rank}}$ as $\tilde{\mathbf{G}}$. Then, a computationally stable rewrite of the MNE estimate is:

$$\hat{\mathbf{X}}^{MNE}\left(\tilde{\mathbf{Y}},\tilde{\mathbf{G}},\mathbf{R}\right) = \mathbf{R}\tilde{\mathbf{G}}'\left(\tilde{\mathbf{G}}\mathbf{R}\tilde{\mathbf{G}}'+\mathbf{I_N}\right)^{-1}\tilde{\mathbf{Y}}, \tag{3.17}$$

where N is the number of sensors. Computing these current estimates requires specification of the prior covariance matrix $\mathbf{R}$. In practice, $\mathbf{R}$ is unknown, and is usually written in terms of a regularization parameter as $\tilde{\mathbf{R}} = \mathbf{R}/\lambda^2$, giving:

$$\hat{\mathbf{X}}^{MNE}\left(\tilde{\mathbf{Y}},\tilde{\mathbf{G}},\tilde{\mathbf{R}}\right) = \tilde{\mathbf{R}}\tilde{\mathbf{G}}'\left(\tilde{\mathbf{G}}\tilde{\mathbf{R}}\tilde{\mathbf{G}}'+\lambda^2\mathbf{I_N}\right)^{-1}\tilde{\mathbf{Y}}. \tag{3.18}$$

We note that this $\hat{\mathbf{X}}^{MNE}$ specifies the currents in modes of the forward solutions, and can be projected back into dipole space for reporting.

The power signal to noise ratio in this whitened case is $\mathrm{SNR} = \mathrm{Tr}\left(\tilde{\mathbf{G}}\tilde{\mathbf{R}}\tilde{\mathbf{G}}'\right)/\mathrm{Tr}\left(\lambda^2\mathbf{I_N}\right)$. Setting $\lambda^2 = 1/\mathrm{SNR}$ and assuming $\tilde{\mathbf{R}}$ is diagonal gives a convenient choice for the elements of $\tilde{\mathbf{R}}$:

$$\tilde{\mathbf{R}}_{ii} = \frac{\mathrm{Tr}\left(\mathbf{I_N}\right)}{\mathrm{Tr}\left(\tilde{\mathbf{G}}\tilde{\mathbf{G}}'\right)}. \tag{3.19}$$

High SNR values enforce source current amplitudes to explain peaks in the data better than the noise. By contrast, low SNR values weight the source covariance more and allow larger discrepancy between measured and predicted data. Standard settings of SNR in the range of $9-25$ were used

for MEG evoked response recordings where the goal is to capture peaks better than background noise. For EEG recordings, as forward solutions have greater blurring due to the relatively low conductivity of the scalp and skull, greater information is contained in background below peaks, and thus SNR was set to 1. The same SNR value was used through the source space hierarchies.

3.4.4 Subspace Pursuit Refinement

The subspace pursuit implementation used to refine and identify a sparse subset of MNE estimates for a given source space was based on [86, 187, 188], stated in algorithm form as SP $(\mathbf{Y},\ \mathbf{G},\ \mathbf{Q},\ \mathbf{R},\ L) \Leftrightarrow$ SP $\left(\tilde{\mathbf{Y}},\ \tilde{\mathbf{G}},\ \tilde{\mathbf{R}},\ L\right)$.

Initialization:

1. Correlation threshold: Compute a correlation threshold μ (termed mutual coherence in [86]) specifying the degree of orthogonality required amongst the selected columns of $\tilde{\mathbf{G}}$. This is set as the average worst case correlation between the forward solutions of neighboring cortical patches in the source space under consideration (§ 3.8.2).
2. Support: Compute the standard MNE estimates $\hat{\mathbf{X}}^{MNE}\left(\tilde{\mathbf{Y}}, \tilde{\mathbf{G}}, \tilde{\mathbf{R}}\right)$ as specified in Eqn. 3.18, and the l_2-norms across time for each row of $\hat{\mathbf{X}}^{MNE}$. Specify initial support $\mathcal{H}^{(0)}$as the set of L rows in $\hat{\mathbf{X}}^{MNE}$ with largest l_2-norms satisfying the degree of orthogonality specified by μ.
3. Residual: $\mathbf{F^{(0)}} = \tilde{\mathbf{Y}} - \tilde{\mathbf{G}}\left(\mathcal{H}^{(0)}\right)\hat{\mathbf{X}}^{MNE}\left(\tilde{\mathbf{Y}},\ \tilde{\mathbf{G}}\left(\mathcal{H}^{(0)}\right),\ \tilde{\mathbf{R}}\right)$.

Iteration starting at $l = 1$:

1. Support Expansion: $\mathcal{H}^{(l)} = \mathcal{H}^{(l-1)} \cup \{L$ rows in $\hat{\mathbf{X}}^{MNE}\left(\mathbf{F}^{(l-1)},\ \tilde{\mathbf{G}}, \tilde{\mathbf{R}}\right)$ with largest l_2-norms satisfying the degree of orthogonality specified by $\mu\}$.
2. Estimation on Expanded Support: $\mathbf{Z}^{(l)} = \hat{\mathbf{X}}^{MNE}\left(\tilde{\mathbf{Y}},\ \tilde{\mathbf{G}}\left(\mathcal{H}^{(l)}\right),\ \tilde{\mathbf{R}}\right)$.
3. Support Trimming: Update $\mathcal{H}^{(l)}$ to L rows in $\mathbf{Z}^{(l)}$ with largest l_2-norms satisfying the degree of orthogonality specified by μ.
4. Residual Update: $\mathbf{F}^{(l)} = \tilde{\mathbf{Y}} - \tilde{\mathbf{G}}\left(\mathcal{H}^{(l)}\right)\hat{\mathbf{X}}^{MNE}\left(\tilde{\mathbf{Y}},\ \tilde{\mathbf{G}}\left(\mathcal{H}^{(l)}\right),\ \tilde{\mathbf{R}}\right)$.
5. Stopping Criterion: If $\mathcal{H}^{(l)} = \mathcal{H}^{(l-1)}$, set outputs $\mathcal{H} = \mathcal{H}^{(l)}$ and $\hat{\mathbf{X}}^{(\text{end})}_{\mathcal{H}} = \hat{\mathbf{X}}^{MNE}\left(\tilde{\mathbf{Y}},\ \tilde{\mathbf{G}}\left(\mathcal{H}^{(l)}\right),\ \tilde{\mathbf{R}}\right)$ and end iterations.

The final $\mathcal{H}$ specifies the subset of regions in the source space that are most relevant to the measurement, while $\hat{\mathbf{X}}^{(\text{end})}_{\mathcal{H}}$ specifies the currents in modes of the forward solutions of these regions (as estimates in Eqn. 3.18 are computed on the reduced order eigenmode approximations of $\tilde{\mathbf{G}}$). Note that the $\hat{\mathbf{X}}^{(\text{end})}_{\mathcal{H}}$ is unitless as the forward solutions are scaled by current strengths (§ 3.2.3). To interpret the regional currents on the same units scale and in physical dipole coordinates, all $\hat{\mathbf{X}}^{(\text{end})}_{\mathcal{H}}$ were

scaled by the current strength c and projected back into dipole space to obtain $\hat{\mathbf{X}}_{\mathcal{H}} = c\mathbf{W}'\hat{\mathbf{X}}_{\mathcal{H}}^{(\text{end})}$. Overall, this subspace pursuit algorithm offers an efficient means to iteratively concentrate the MNE into the few regions most "relevant" (in the minimum l_2-norm sense) to the measurement. This algorithm is repeated across hierarchies, to yield estimates $\hat{\mathbf{X}}_{\mathcal{H}_{\mathbb{J}}}$ at the final hierarchy, denoted as $\hat{\mathbf{X}}$ for notational convenience.

3.4.5 Hierarchical Data Analysis

Forward solutions were computed across the hierarchies of source spaces and grouped by region. All cortical analyses were run to $\mathbb{C}^{(3)}$. Time windows of interest were picked based on latencies of expected features. The number of relevant regions or the target sparsity level L was specified to reflect the spatial extent/number of regions expected to be active during a given paradigm. The same L was maintained across hierarchies. While the number of relevant regions was pre-specified for convenience, this is not essential as the algorithm could be rewritten to iteratively add one projection at a time until all significant field patterns in the measurement are explained and the residual is white across sensors. This procedure would be very similar to employing a formal model selection and goodness of fit procedure. Both MEG and EEG recordings were used for localization if simultaneous recordings were available. The two sets of recordings were processed separately (filtering, averaging and covariance estimation), their forward solutions were computed separately, and the measurement vectors and forward solutions were concatenated after accounting for respective noise covariances separately for consistency of units.

All visualizations include the MNE estimate on the final joint source space $\mathbb{J}$ alongside the subspace pursuit estimate in $\mathbb{J}$, to show how SP concentrates the MNE estimate into the most relevant projections. All estimated source current estimates (for MNE and SP) are displayed as resultant currents by region. Resultant current for a given region is defined as the vector sum of source currents across all dipoles and orientations in that region. Thus, the results presented are not sensitive to the specific spacing between dipoles. Spatial maps are shown on inflated surfaces for the cortical stages and representative MRI slices for the final joint stages. Full regional distributions of source currents are summarized in bar graphs as follows. Denoting $\hat{\mathbf{x}}$ as the current time course for a dipole in $\hat{\mathbf{X}}$ (1 row of $\hat{\mathbf{X}}$), the root mean square (RMS) magnitude of dipole current across time is summarized by $\sqrt{||\hat{\mathbf{x}}||_2}$. Taking the mean of these values across dipoles within an anatomic region, we obtain the RMS magnitude of current in a representative dipole for each region. Computed across regions, this gives a quantitative picture of regional distribution of current estimates.

3.5 Estimation Results

3.5.1 Somatosensory Evoked Response Simulations

To illustrate the algorithm, we first analyze simulated MEG evoked responses corresponding to a somatosensory right median nerve stimulation paradigm. These evoked responses progress from

right primary median nerve afferent, up the spinal cord and brainstem, to a tertiary neuron in left ventral posterolateral nucleus of thalamus (VPL), then to the left somatosensory cortices near the post central gyrus, and later to the left posterior parietal cortex and bilateral second somatosensory cortices. The VPL thalamic peaks typically have low amplitudes, and thus offer a challenging but realistic test case.

3.5.1.1 Simulation Design

We obtained anatomic MRI on a healthy volunteer, constructed distributed cortical and subcortical source spaces, and computed corresponding electromagnetic forward models. We simulated dipole activity within 5 regions of interest - namely a $1\,\mathrm{cm}^3$ volume in left VPL thalamus, $\sim 600-800\,\mathrm{mm}^2$ surface patches spanning primary and secondary somatosensory cortices and the posterior parietal area (Fig. 3-5). The time courses of the simulated currents were a series of Gabor atoms of the form $Ae^{-(t-t_o)/2\sigma^2}$ where A, t_o and σ denote the amplitude (based on regional current strengths and amplitude ranges from [86, 119, 190]), delay and width of the evoked component respectively. Currents were generated over a time range of $0-220\,\mathrm{msec}$, with $3\,\mathrm{kHz}$ sampling rate (Fig. 3-5A). VPL activity is brief and occurs before any cortical regions are activated to reflect the physiological sequence. The resulting field patterns at the sensors are obtained by putting dipole currents for each region of interest through the corresponding electromagnetic forward models, and summing the contributions from all regions (Fig. 3-5B). As expected, VPL activation generates significantly lower MEG amplitudes than cortical activations. White zero mean Gaussian noise was added on to these field patterns for a realistic SNR of $\sim 7\,\mathrm{dB}$ (Fig. 3-5C); this puts VPL activity well under the noise floor. VPL amplitudes, dipole locations and source current shapes were varied for a total of 30 simulations.

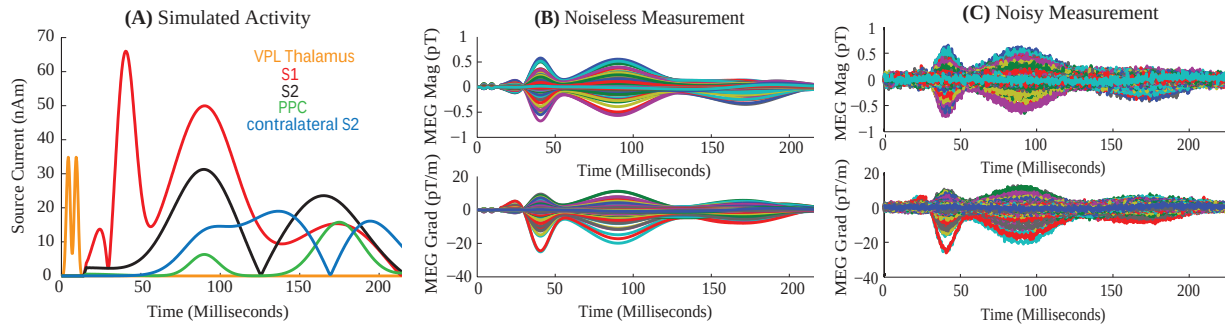

Figure 3-5: Simulated Somatosensory Evoked Responses. (A) Simulated source currents. The VPL thalamus is activated during the first 15 msec of the evoked response, followed by activity in primary and secondary somatosensory cortices, and the posterior parietal area. *(B)* MEG measurements due to the source currents in Panel A, projected to sensors via the electromagnetic forward model. The fields due to activity in VPL are significantly smaller than the cortically generated fields. *(C)* Gaussian noise is added on to the measurements in Panel B, driving fields arising from VPL well below noise floor.

3.5.1.2 Localization Results

To test whether it is possible to resolve specific thalamic subdivisions using non-invasive recordings, we applied our hierarchical subspace pursuit algorithm to the simulated somatosensory data. The input parameter L for this case can be varied from $2-8$, as we expect activity in $2-4$ distinct regions at any given time, and multiple modes can be allowed per region. The following figures show results for the most flexible case of $L=8$.

Fig. 3-6 illustrates the hierarchical reduction of cortical source spaces. Fig. 3-6*A* shows spatial maps of the simulated data, and Fig. 3-6*B-D* show subspace pursuit estimates progressing through source spaces of increasing fineness. While the coarse cortical solutions generally localize to appropriate areas, they also contain incorrect components, e.g., left ventral and right auditory components, but these irrelevant components do not appear in finer cortical source spaces. Further, the spatial resolution increases through finer source spaces, giving sparse cortical solutions in $\mathbb{C}^{(3)}$ with late components in somatosensory and parietal cortices accurately matching simulations.

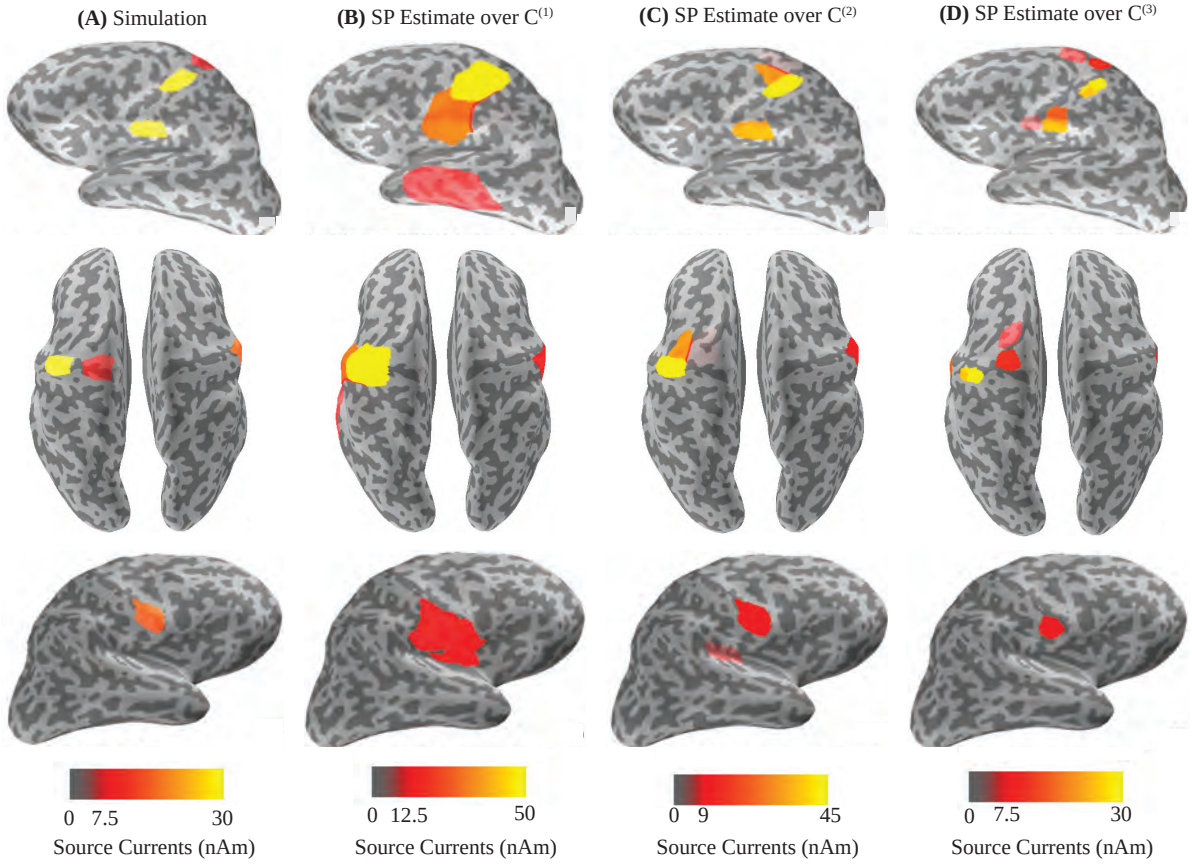

Figure 3-6: Hierarchical Reduction of Cortical Source Space Relevant to Simulated Data. (*A*) Simulated cortical activity mapped to the inflated surface. *(B-D)* Cortical maps of subspace pursuit estimates, showing the systematic refinement across the progressively finer source spaces $\mathbb{C}^{(1)}$, $\mathbb{C}^{(2)}$, and $\mathbb{C}^{(3)}$. Estimates displayed are resultants across dipoles within their respective patches. All topographical snapshots are at 90 msec (simulated time courses in Fig. 3-5*A*).

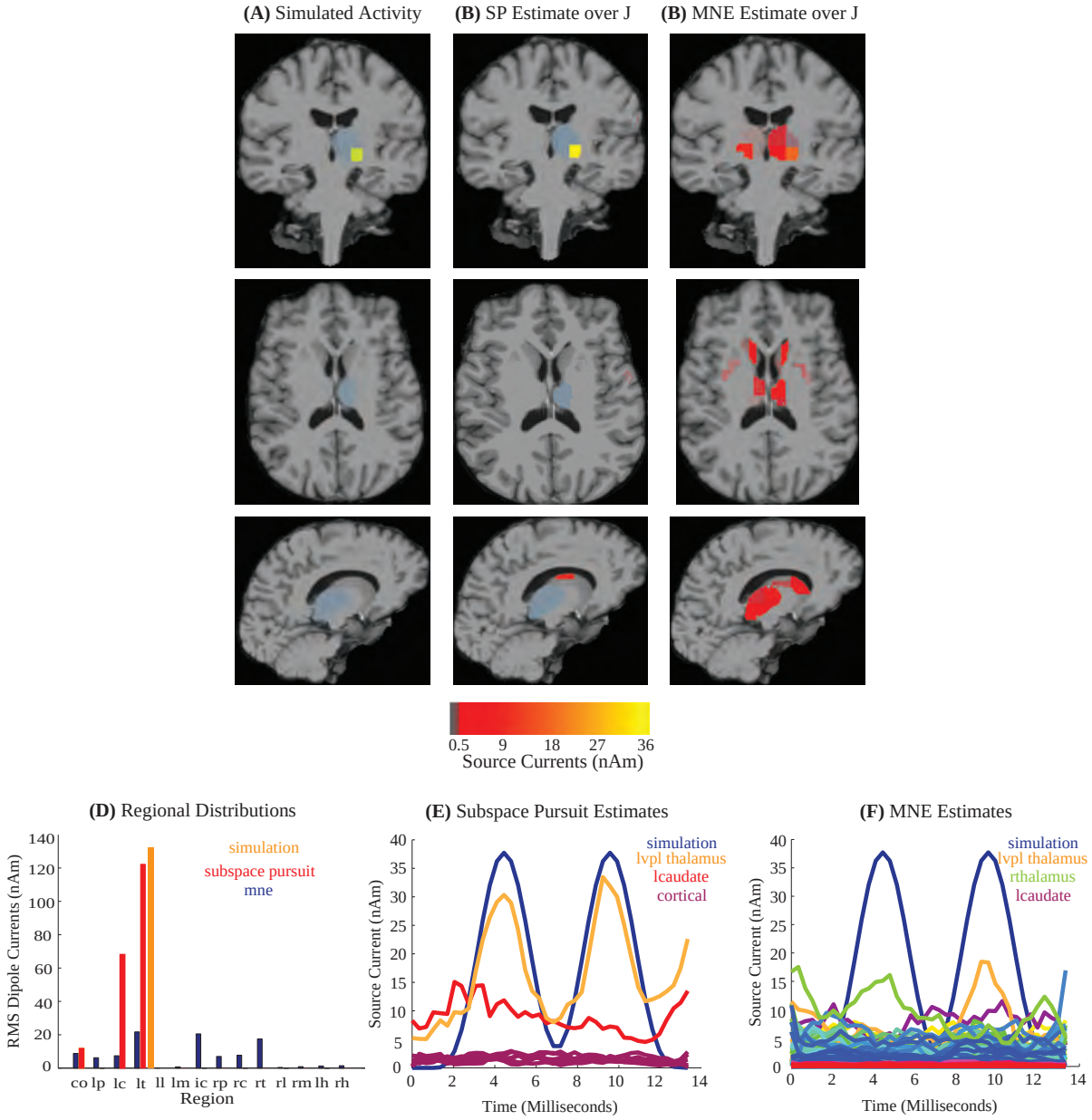

Figure 3-7: Subcortical Estimates Explain Simulation. *(A-C)* Representative MRI maps showing regions with simulated activity and the corresponding subspace pursuit (SP), and MNE estimates obtained upon the reduced source space $[\mathbb{C}_{sp}, \mathbb{S}]$. The translucent blue region is left thalamus shown for reference only. MRI slice locations in the surface RAS coordinate system are 125 mm (coronal), 120 mm (horizontal) and 137 mm (sagital). *(B vs. C)* Subspace pursuit estimates are more specific to the simulated region than MNE estimates. *(D)* Summary of mean activity by anatomic region for the simulation and two estimation algorithms. Legend: cortical (co), putamen (r/lp), caudate (r/lc), thalamus (r/lt), hippocampus (r/lh), lateral geniculate (r/ll), medial geniculate (r/lm), and inferior colliculus (ic), where r/l refer to right/left. *(E-F)*. Time courses of source current estimates from subspace pursuit and MNE algorithms, overlaid against simulation in VPL thalamus (simulated time courses in Fig. 3-5*A*). All estimates displayed are resultants across dipoles within their respective patch or volume subdivisions. Topographical snapshots are shown at 9 msec. Each time course shown corresponds to a selected subdivision in the respective anatomic regions indicated in legends.

These cortical solutions, denoted as $\mathbb{C}_{sp}$, along with the subcortical volume sources $\mathbb{S}$, form the final combined source space $\mathbb{J}$ for final estimation. We focus on describing estimates for the early 15 msec section where VPL is activated. Fig. 3-7 illustrates spatial maps and time courses of the estimates within source space $\mathbb{J}$. The true simulated activity is shown in Fig. 3-7*A*. SP estimates, in Fig. 3-7*B*, have a high amplitude component in VPL thalamus, resembling the true simulated activity in spatial configuration. There is a small subdivision in caudate (seen in sagittal slice) which also captures some energy, but most energy is concentrated into VPL.

Fig. 3-7*C* shows the MNE estimate on $\mathbb{J}$ for comparison. While the MNE estimate has a hotspot in VPL thalamus, it is also smeared out widely to regions in other parts of thalamus and caudate. To quantify this comparison, we inspected regional distributions of dipole current amplitudes by anatomic region, as summarized in Fig. 3-7*D*, The SP estimate is concentrated in left thalamus, while the MNE estimate in left thalamus is comparable to inferior colliculus and right thalamus, with some additional components in basal ganglia. Thus, it is clear that SP refines the MNE estimate by concentrating current into the few regions/projections most relevant to the measurement.

Next, we inspected time courses of the SP and MNE estimates in Fig. 3-7*E-F*. For the SP estimates, the VPL thalamus captures the shape of the simulated current, while the estimate in caudate has spurious shape and all other regions have near null activity. The shape of the caudate appears much like an offset. Therefore, we asked if the caudate enters the solution primarily to cancel out the net effects of the many near null components. Indeed, dropping the target sparsity level from $L = 8$, to $L = 3$ takes away the component in caudate and makes the SP solution specific to VPL thalamus with near null components in all other regions, validating the specificity of our algorithm (Fig. 3-12). For the MNE estimates, the shape of the simulated current is seen across thalamus, caudate and inferior colliculus, with amplitudes split between these regions, making it hard to discern specific active regions.

Finally, we note that these localization results are robust despite 20X drop in amplitude of simulated currents within VPL, suggesting that the algorithm localizes sources based on their field patterns and is not sensitive to the amplitude of fields arising from subcortical sources.

3.5.2 Auditory Evoked Response Recordings

To validate the algorithm further, we chose to analyze auditory responses evoked with a train of click stimuli during resting eyes open condition. Auditory responses comprise distinct M/EEG peaks and established latencies corresponding to a stereotypical progression of activity from the cochlea, through inferior colliculus to medial geniculate thalamus and auditory cortex [191], and thus serve as a good test case for validating a subcortical source localization algorithm. We obtained simultaneous M/EEG auditory evoked recordings (AEPs) during binaural stimulation and corresponding anatomic MRI using the paradigm described below.

3.5.2.1 Paradigm and Study Design

Simultaneous MEG and EEG AEP recordings, and structural MRIs were obtained on two healthy volunteers aged $25-45$ years screened for standard MRI contraindications and normal audiometry (no evidence of hearing loss in $0-4$ kHz range). The paradigm design was based on [117, 192]. A train of broadband clicks having 0.1 msec duration, intensity $65-80$ dB/nHL and inter-stimulus interval 110 msec (click rate 9.09 Hz, corresponding to highest AEP SNR), generated within the Presentation^{TM} software (Version 17.1, Neurobehavioral Systems, Inc., Albany, CA, USA), was delivered binaurally during eyes open resting condition. Subjects were asked to sit still, not pay any attention to the sounds and imagine a dot at the center of the screen. M/EEG data were recorded at 5 kHz sampling with filter cutoffs set to $0.03-1660$ Hz. After acquiring 2 min of pre-stimulus baseline eyes open data, AEPs were recorded in 5 runs of 5.5 min each, yielding $10000-16000$ epochs for averaging. At the start of each run, standard checks were performed to ensure no stimulus artifacts, subjects were allowed a blink break if needed, and HPI recordings were obtained. Non-stimulus baseline recordings were repeated at the end of the study and MEG emptyroom recordings were also obtained. The study was approved by the Partners Human Research Committee at Massachusetts General Hospital (protocol number 1999-P-010946), and written informed consent was obtained from all subjects.

3.5.2.2 Data Processing

The raw data were preprocessed to remove power line noise using a comb notch filter (MATLAB^{TM}) with 30 notches at harmonics of 60 Hz, each having bandwidth 1 Hz. Artifactual channels (marked by inspection) and eye-blink epochs (peak to peak EOG<150 uV in $1-40$ Hz band) were both excluded. As is standard in the AEP literature [117, 191, 192], the early auditory brainstem reponse (ABR) and later middle latency response (MLR) components of the auditory evoked potential were processed separately by band-pass filtering the preprocessed data to $500-1625$ Hz, and $30-300$ Hz respectively. In each case, the data were processed to (a) compensate for interference, (b) obtain stimulus locked average evoked responses, and (c) estimate noise covariances.

First, standard signal space projections (SSP), computed using principal component analysis on the empty room MEG recordings, were used to compensate for environmental interference [117, 189]. Second, stimulus timings to denote epochs for averaging were compensated for a sound tube delay of 9.775 msec (as measured by recording sounds delivered at the ear piece with an optical microphone). Grand averages across runs were performed without HPI correction if HPI head coordinates across runs were within 2 %. Overall, a total of 11200 epochs were used for the averages considered in our analysis. Third, observation noise covariances were estimated using the baseline eyes open recording. These data give a good measure of background "noise" arising from ongoing brain activity and systematic instrumental disturbances [189]. As the MEG noise covariance is typically ill-conditioned, external disturbances to the covariance estimate were suppressed by applying the SSP operator, and then diagonal loading was applied ([189], § 3.8.3). All SSP, averaging and

covariance calculations were performed in bands relevant to components of interest.

3.5.2.3 Data Summary

Fig. 3-8 summarizes the M/EEG evoked responses. The early auditory brainstem responses (ABR) are shown in Fig. 3-8*A*, while later middle latency responses (MLR) are in Fig. 3-8*B*.

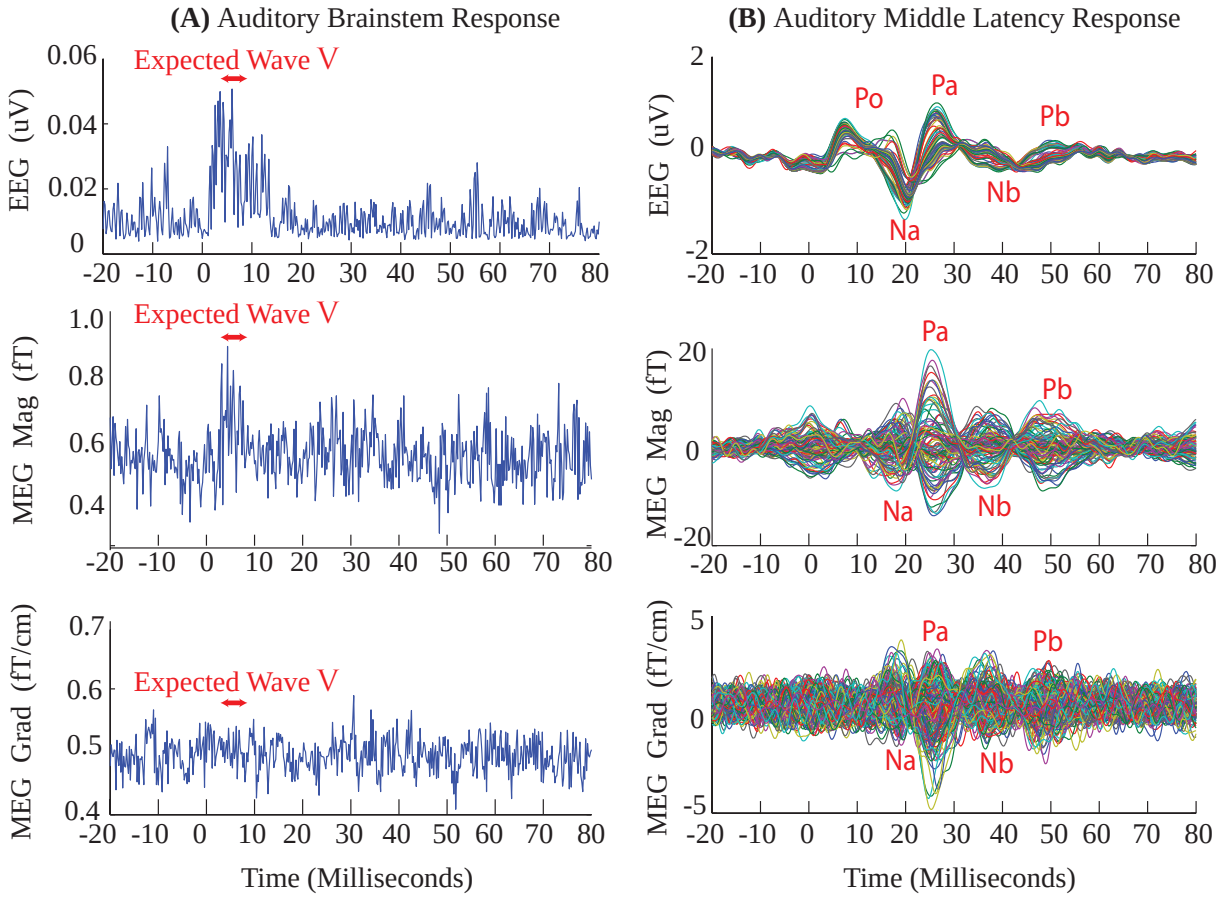

Figure 3-8: Auditory Evoked Response Recordings. Stimulus locked average auditory evoked responses obtained on a healthy volunteer presented with a broadband click train stimulus. Data shown are averages across 11200 epochs. (*A*) Auditory brainstem responses - the red bar marks periods when early peaks are seen in the data, the timing of these peaks is consistent with expected location of wave V. Time courses are raw recordings filtered $500-1625$ Hz, rectified and averaged across channels. *(B)* Middle latency responses. Common peaks are marked and latencies follow known trends. Time courses are raw recordings filtered $0-300$ Hz, displayed across multiple channels.

Examining Fig. 3-8*A*, we note that the EEG ABRs have higher SNR than MEG ABRs. Thus marking the early peaks on EEG and looking for coincidental peaks in the MEG recordings, we see that the MEG magnetometers also show ABRs at the expected latencies while the gradiometers do not exhibit significant features at these latencies. While it is difficult to resolve individual early waves corresponding to cochlear nucleus and lower brainstem structures like the superior olivary complex, there is clear evidence of evoked peaks in the $5-7$ msec time range, which is consistent

with the expected timing ($5.6 - 5.8$ msec) of the brainstem wave V known to arise from inferior colliculus [117, 191]. Proceeding to the MLR in Fig. 3-8*B*, we see that the EEG has a low amplitude Po feature, which is thought to mark the end of brainstem components and potentially the start of thalamic components [191]. However, Po is not seen clearly on the MEG and the thalamic involvement at these latencies is controversial and has not been resolved to date [117, 191]. Peaks Na-Pa reflect classic middle latency activity in the (supratemporal) primary auditory cortex (A1). Peaks beyond 40 msec reflect activity in secondary auditory cortices and higher association regions. Thus, as the most prominent peaks with well-known locations are the early brainstem wave V and the later Na-Pa peaks, we asked if source localization can reveal activity in (a) inferior colliculus during $5 - 7$ msec, and (b) A1 during $20 - 30$ msec.

3.5.2.4 Localization Results

Fig. 3-9 illustrates the spatial maps and time courses of source estimates obtained on the auditory evoked responses. Fig. 3-9*A* shows localization of MLR peaks (Na-Pa) to bilateral primary auditory cortices (A1). These areas comprise the reduced cortical source space $\mathbb{C}_{\text{sp}}$, which along with the subcortical volume sources $\mathbb{S}$, form the combined source space $\mathbb{J}$ for final estimation. We focus on describing estimates for the early ABR components where subcortical structures are activated. Fig. 3-9*B-D* illustrates spatial maps and time courses of the estimates within source space $\mathbb{J}$. Subspace pursuit estimates (Fig. 3-9*B)* localize specifically to inferior colliculus, while MNE estimates (Fig. 3-9*C)* localize both to inferior colliculus and sources in caudate and thalamus (with the latter estimates having higher amplitude than the former). Fig. 3-9*D* quantitatively summarizes the regional distributions of dipole current amplitudes by region - showing that SP refines the non-specific MNE estimate to concentrate activity into the most relevant regions.

Time courses of the SP and MNE estimates are in Fig. 3-9*E-F*. SP estimates of activity in inferior colliculus peak at the time interval where wave 5 is expected. We note that as the number of relevant regions L was set to 1 (only 1 key region is expected to be active during $0 - 10$ msec), the sparse estimate shows only one time course. The MNE estimates show activity in multiple regions peaking at the time intervals of interest, making it hard to discern specific activity expected in inferior colliculus.

Finally, as all the above estimates were obtained using both MEG and EEG data (where EEG has higher SNR wave 5 peaks), we tested efficacy on unimodal MEG data, which has lower SNR. We found that subspace pursuit results on MEG data are just as accurate and specific as the estimates on MEG-EEG data - thus showing that the algorithm does not rely on having high amplitude peaks or high SNR in the data.

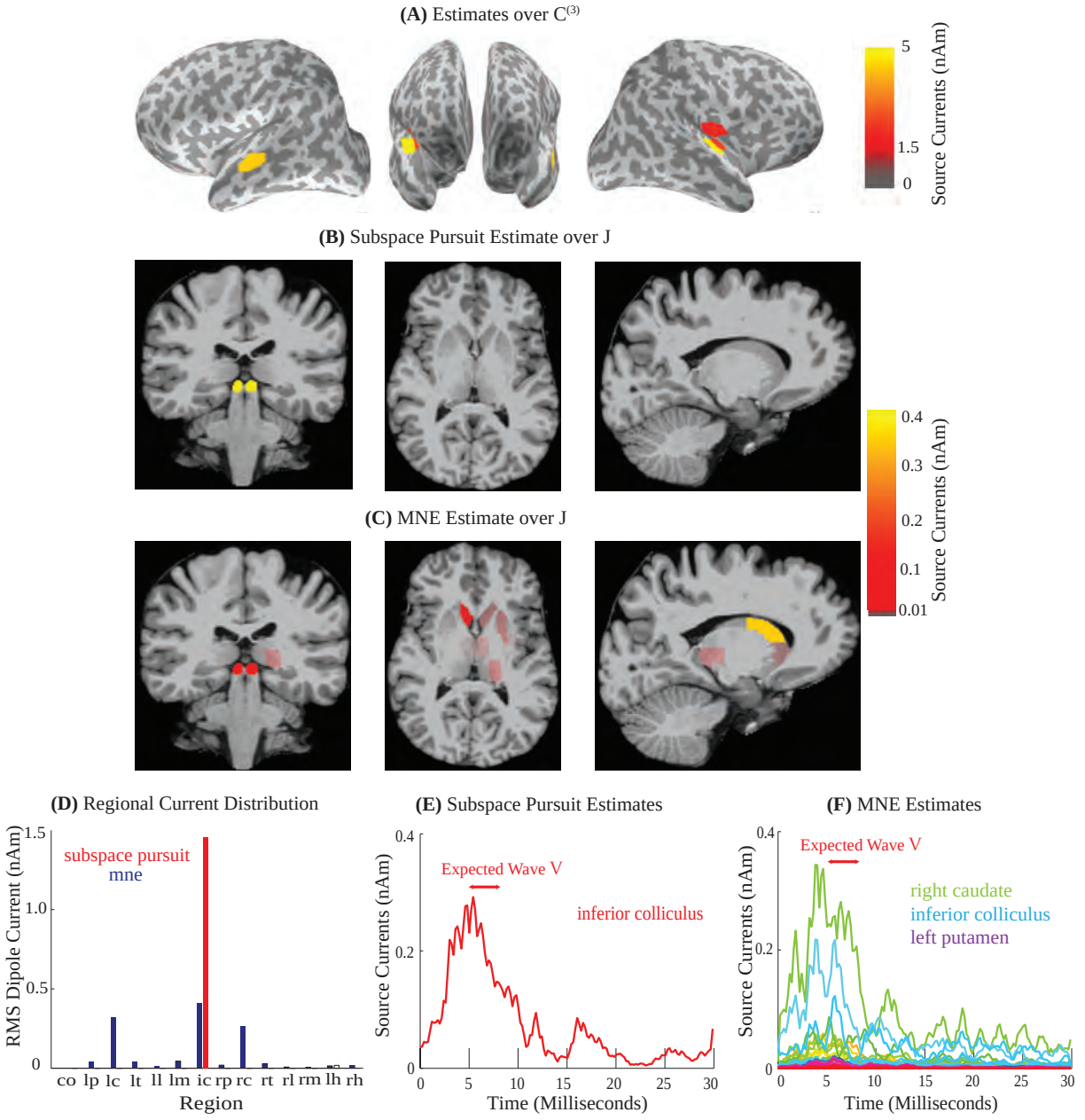

Figure 3-9: Cortical and Subcortical Estimates Explain Auditory Brainstem Recordings. (*A*) Reduced cortical source space obtained as a result of initial source space reduction performed upon the middle latency recordings. Estimates shown on the inflated surface are snapshots at 25 msec (data traces in Fig. 3-8*B*) and thus comprise patches in primary auditory areas. SP and MNE algorithms were then applied on the reduced source space $[\mathbb{C}_{\text{sp}}, \mathbb{S}]$, to localize sources of the auditory brainstem recordings. *(B-C)* Representative MRI maps showing snapshots of subspace pursuit and MNE estimates at 5 msec (data traces in Fig. 3-8*A*). MRI slice locations in the surface RAS coordinate system are 115 mm (coronal), 126 mm (horizontal) and 142 mm(sagital). *(D)* Summary of mean activity by anatomic region for each estimation algorithm. Legend: cortical (co), putamen (r/lp), caudate (r/lc), thalamus (r/lt), hippocampus (r/lh), lateral geniculate (r/ll), medial geniculate (r/lm), and inferior colliculus (ic), where r/l refer to right/left. *(E-F)* Time courses of source current estimates obtained with SP and MNE algorithms. *(D-F)* SP is more specific to the inferior colliculus than MNE. All estimates displayed are resultants across dipoles within their respective patch or volume subdivisions. Each time course shown corresponds to a selected subdivision in the respective anatomic regions indicated in legends.

3.6 Performance Analysis

3.6.1 Hierarchical Reduction is Essential to Resolve Subcortical Activity

All SP and MNE estimates described above are performed after a hierarchical reduction of the cortical source space. To better understand how this reduction affects the solution, we performed subspace pursuit and MNE for the simulation study on sources across the brain in a single stage i.e., on source space $[\mathbb{C}^{(3)}, \mathbb{S}]$. This estimation converges only on cortical regions (Fig. 3-10*A*). Regional distributions of estimated activity for early time points show that neither MNE nor SP can capture the simulated activity: the SP solution is entirely cortical, while the naive MNE solution is barely discernable. Hence, a single stage solution, will, in general converge to cortical areas even when subcortical sources are active.

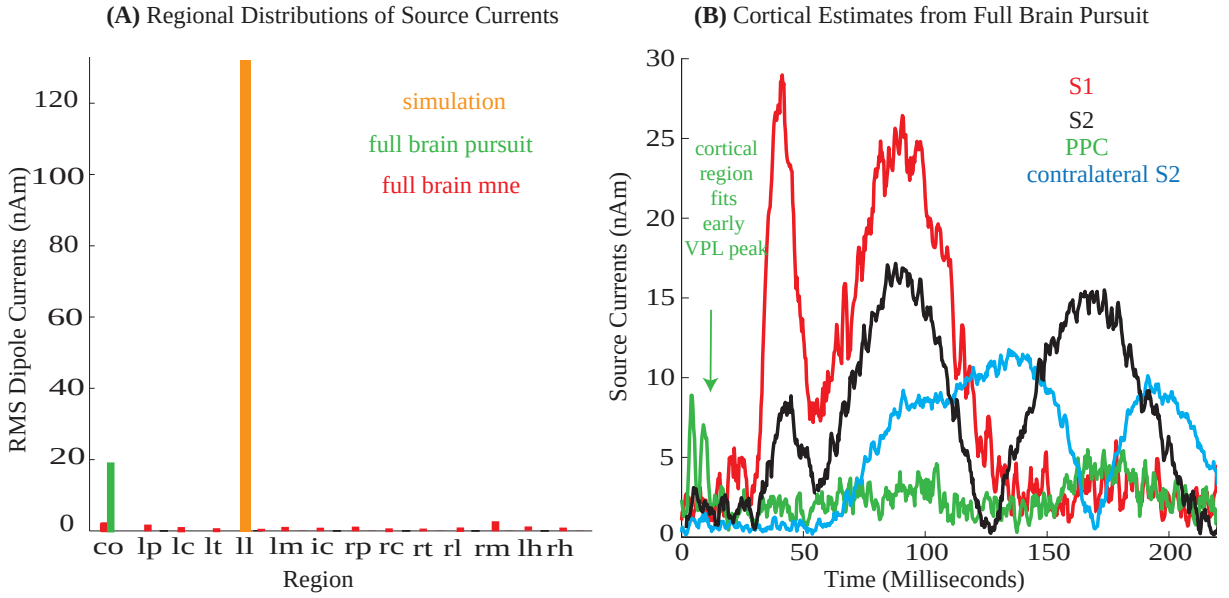

Figure 3-10: Subcortical Activity Cannot be Resolved without Hierarchical Reduction of Source Spaces. (*A*) Summary of mean activity by anatomic region for the simulation (in Fig. 3-5) and different single stage estimation algorithms (estimation performed on full brain source space $[\mathbb{C}^{(3)}, \mathbb{S}]$). Legend: cortical (co), putamen (r/lp), caudate (r/lc), thalamus (r/lt), hippocampus (r/lh), lateral geniculate (r/ll), medial geniculate (r/lm), and inferior colliculus (ic), where r/l refer to right/left. All SP estimates are in cortex, and MNE estimates have very low amplitude spread across the brain. *(B)* Time courses of cortical source current estimates obtained with the SP algorithm. Early VPL currents are explained by cortical regions (purple arrow). All estimates displayed are resultants across dipoles within their respective patches (simulated time courses in Fig. 3-5*A*). Each time course is for a selected patch in the respective anatomic regions indicated in legends.

We further examined the SP source current estimates, all within cortex, over time (Fig. 3-10*B*). Cortical currents during the early period resemble the simulated VPL current shape (seen in Fig. 3-5*A*), suggesting that the cortical regions can explain the subcortical simulation. This

suggests that typical cortical solutions, which do not properly account for subcortical sources, may contain spurious activity arising from subcortical regions.

Overall, these observations are consistent with our field comparison analysis and further illustrate our insights that resolving subcortical sources, amidst cortical sources, requires a hierarchical reduction of the possible source spaces across the brain. Further, they show that purely cortical solutions may have misleading components arising from subcortical sources.

3.6.2 Remarks on Sensitivity to Data Analysis Choices

It is noteworthy that the performance of the subcortical subspace pursuit algorithm is in general insensitive to specific data analysis choices.

First, the estimates are relatively insensitive to the size of patches in the reduced cortical source space $\mathbb{C}_{\text{sp}}$, as principal angle profiles distinguishing subcortical and cortical sources are similar across cortical patch size (Fig. 3-11). Thus, as long as the cortical sources are in accurate regions, the specific patch size does not much matter.

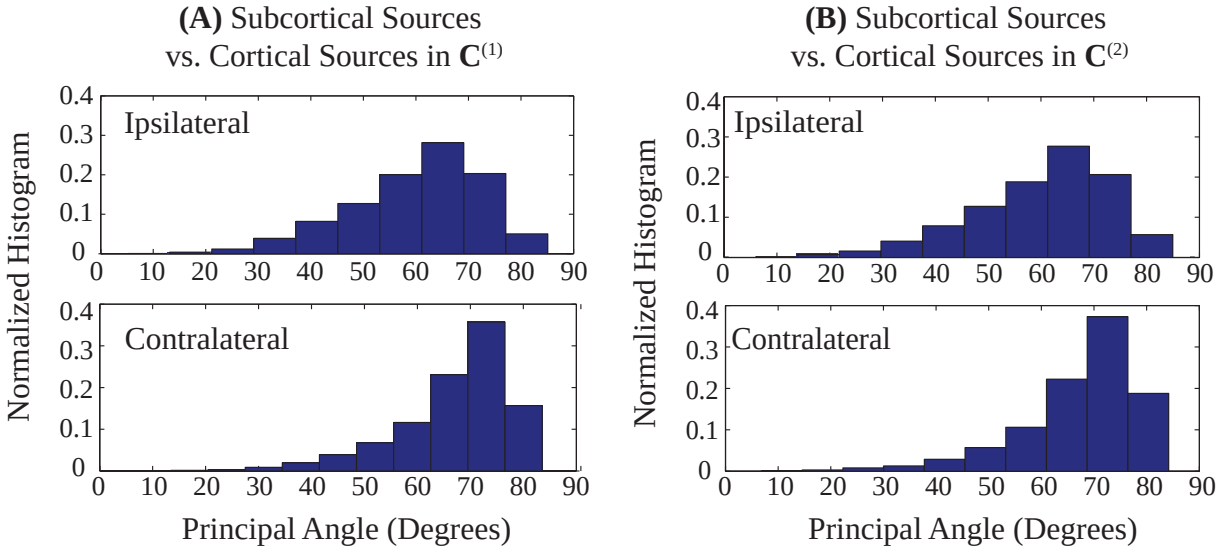

Figure 3-11: Subcortical Estimation Accuracy is Relatively Insensitive to Size of Cortical Patches in Reduced Source Space. Normalized histograms of principal angles quantifying distinctions between MEG field patterns from subcortical and cortical regions. Principal angle profiles for *(A)* coarse subdivisions $\mathbb{C}^{(1)}$ with average patch area 2500 mm^2 , and *(B)* finer subdivisions $\mathbb{C}^{(2)}$ with average patch area 650 mm^3 . Comparable profiles for finer subdivisions $\mathbb{C}^{(3)}$ with patch areas 175 mm^2 are in Fig. 3-2*B*. The field patterns are just as distinguishable across patch sizes, suggesting that accuracy of subcortical estimation is insensitive to size of cortical patches in the reduced source space $\mathbb{C}_{\text{sp}}$.

Second, the estimates are not very sensitive to the specific choice of L within a general range (Fig. 3-12). For very large values of $L > 15$, mirror sources start to appear, but a model selection procedure can provide reasonable ranges for L that appropriately explain the data.

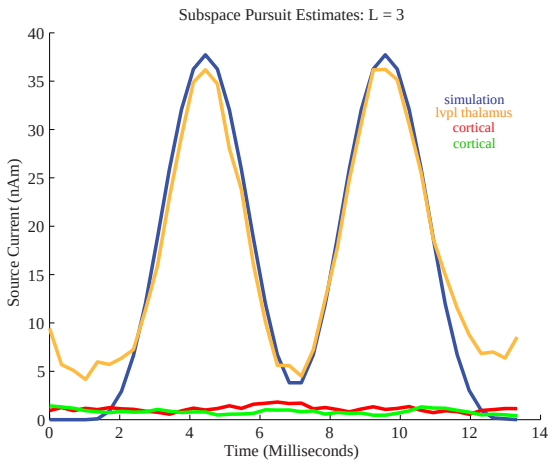

Figure 3-12: Subcortical Estimation Accuracy is Relatively Insensitive to Target Spasrity Level Parameter. Time courses of subcortical source current estimates obtained with SP algorithm for target sparsity level set to $L = 3$ (same case as in Fig. 3-7*E* which displays estimates for $L = 8$). All estimates displayed are resultants across dipoles within their respective patch or volume subdivisions. Each time course is for a selected patch in the respective anatomic regions indicated in legends.

However, we note that this subcortical estimation algorithm is sensitive to two factors.

First, the algorithm can be generally sensitive to the type of measurement employed in estimation. We find that using only EEG data to estimate subcortical sources may not perform as well as MEG data. Fig. 3-13 summarizes the results of the EEG field map analysis. Comparing these results with the MEG field maps in Fig. 3-2, we see that field maps arising from subcortical and cortical sources are less distinct for EEG than for MEG. This may be because of the distinct biophysics underlying generation of EEG and MEG measurements. It may also be because EEG signals are spatially distorted or "blurred" because the skull and scalp have lower conductivity than brain and cerebrospinal fluid. We note that this is a counter-intuitive result: solutions of Maxwells equations suggest that EEG signal amplitudes attenuate less rapidly with increasing source depth than do MEG signal amplitudes [6]. However, even though EEG may record higher SNR signals from subcortical sources, the amplitude-insensitive EEG field maps arising from subcortical and cortical sources are less distinguishable than are the MEG field maps from these sources. Thus, our results counter the typical belief that EEG is better suited than MEG for subcortical source imaging [6, 9]. Further, this analysis suggests that simultaneous M/EEG, combining the high SNR of EEG with the better field map distinctions of MEG, may be best suited for subcortical source estimation.

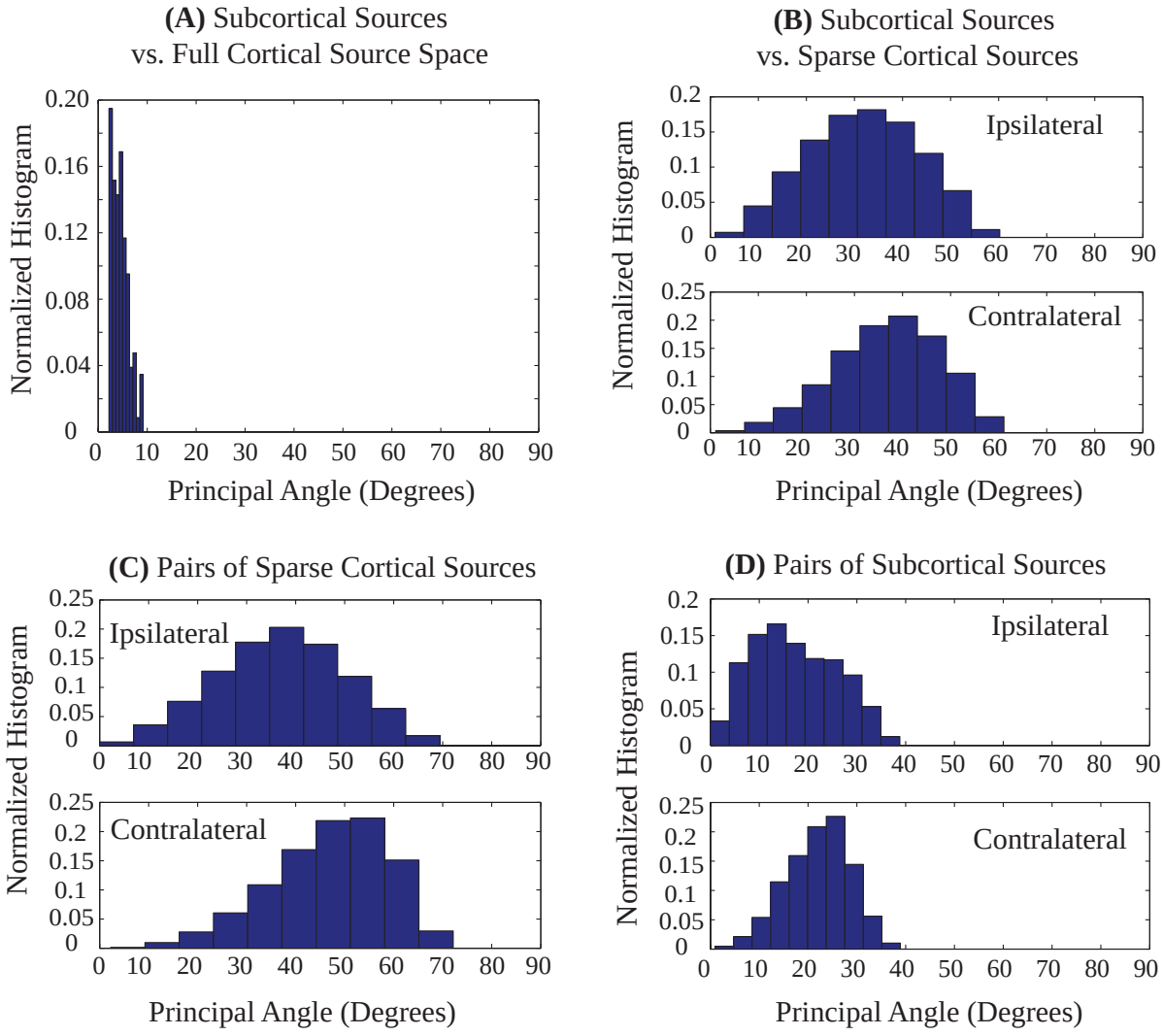

Figure 3-13: Subcortical Estimation Accuracy is Sensitive to Type of Measurement Used, as EEG Field Patterns are More Correlated than MEG Field Patterns. Normalized histograms of principal angles between pairs of brain regions for EEG field patterns. Angles quantify distinctions between subspaces defining all field patterns from the two regions. The general trends across Panels *A*-*D* illustrated in Fig. 3-2 are intact: principal angles are higher when sparse regions are compared (Panels *B*-*D*) as opposed to extensive regions like in Panel *A*. However, EEG field patterns are less distinct across the board than MEG field patterns illustrated in Fig. 3-2. All field pattern comparison results here are shown for $\mathbb{S}$ and $\mathbb{C}^{(3)}$ but hold for coarser cortical source spaces too.

Second, the ability of the algorithm to localize to small nuclei is sensitive to the source space construction, and in particular, to the subdivisions of subcortical volumes. This is illustrated in Fig. 3-14, with the ABR test case. If the source spaces are constructed such that the base subdivision has volume $800\,\mathrm{mm}^3$ (§ 3.8.1), the subspace pursuit finds sources in inferior colliculus as well as hippocampus. Interestingly, the hippocampal sources are diametrically positioned around the tiny medial geniculate nuclei (with volume $150\,\mathrm{mm}^3$). Thus, it appears that the algorithm confounds

sources in the medial geniculate nuclei with larger but diametrically opposite hippocampal sources that can produce the same field pattern as the medial geniculate nuclei. Though the thalamic involvement at these latencies is controversial and has not been resolved to date [117, 191], this result suggests that there may be some sources in the medial geniculate thalamus that may not be easily localized. We note that this non-specificity in localization comes about primarily because subcortical sources have highly inhomogenous volumes and current densities, and can be addressed with improved methods to subdivide subcortical source spaces.

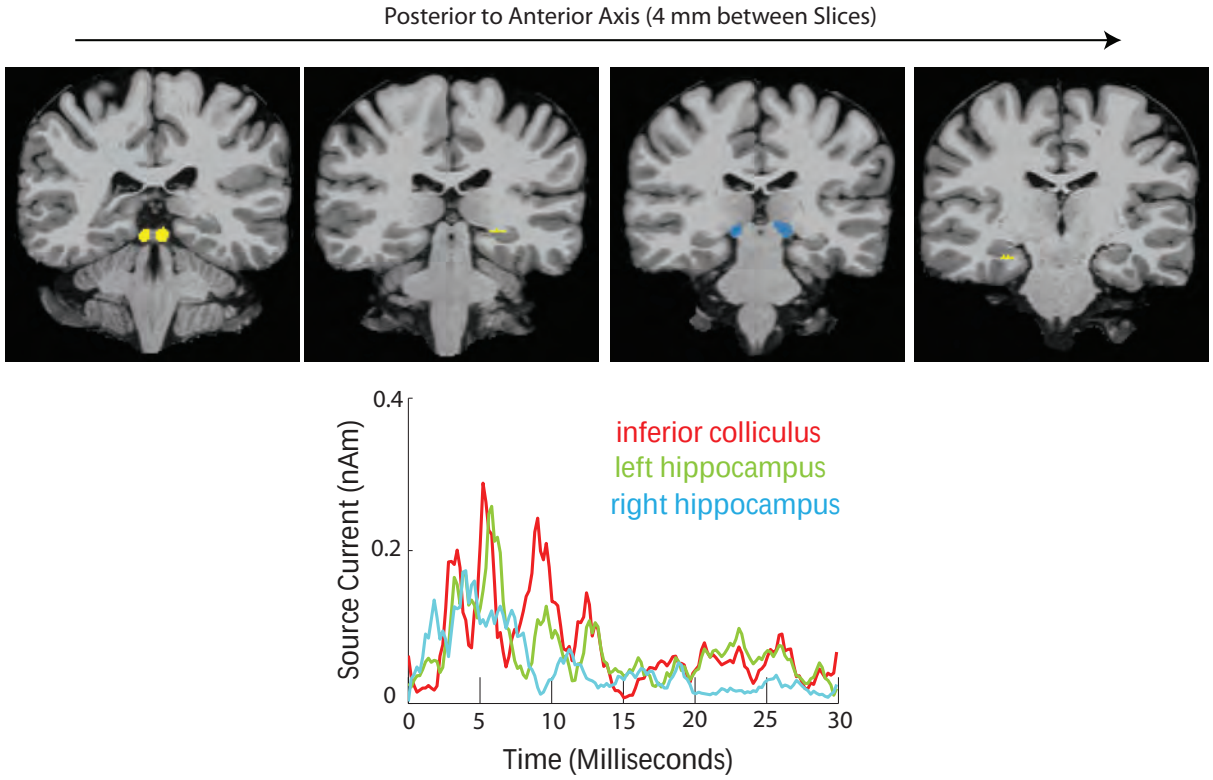

Figure 3-14: Subcortical Estimation Accuracy for Small Nuclei is Limited by Source Space Construction. Source spaces with base subdivision significantly larger than the nuclei of interest, in this case medial geniculate nuclei, lead to solutions with predictable confounds. In the ABR study, the ability to capture medial geniculate nuclei is limited, among other things, by the fact that larger diametrically opposite sources can produce similar field patterns as smaller central sources. As larger sources have greater current strengths, they preferentially enter the solution in lieu of the smaller sources, causing a predictable and easily explained confound. *(A)* Spatial maps showing regions of estimated activity in yellow and potential ground truth regions in blue. Yellow overlays blue for infereior colliculus in the left-most panel. Estimates are in inferior colliculus as well as hippocampal patches diametrically opposite to potential sources in medial geniculate thalamus. *(B)* Hippocampal sources identified have current time courses similar to the time course in inferior colliculus, and thus may correspond to true medial geniculate contributions.

Beyond the above specific points, accurate estimation of noise covariances is essential for good performance of the algorithm. Further, selection of time points of interest can affect performance, as the subspace pursuit refines $\hat{\mathbf{X}}^{MNE}$ based on its norm within specified time intervals. But these points apply just as well to cortical source estimation algorithms as they do to our algorithm.

3.7 Discussions

Subcortical sources are thought to be difficult to localize with M/EEG because they contribute low amplitude (gain) signals to scalp fields. We have identified gain-insensitive information in the M/EEG measurements that allows robust distinctions between (and among) subcortical and cortical sources. Specifically, these distinctions are found in field patterns arising from spatially sparse subcortical and cortical sources. We then translated these insights to the practical problem of localizing subcortical sources. Specifically, we developed a subspace pursuit based procedure that hierarchically narrows the space of possible sources to sparse subsets whose field patterns match those in the measurement. Finally, we applied our approach and algorithm on challenging simulated and experimental evoked response test cases to validate efficacy for localizing non-invasive M/EEG measurements arising from activity in specific thalamic and brainstem nuclei. Taken together, these results demonstrate feasibility of resolving electromagnetic activity within subcortical structures using non-invasive M/EEG recordings.

3.7.1 Field Pattern Distinctions and Sparsity

It is widely believed that M/EEG measurements contain little information to resolve subcortical activity. Our key novel finding is that there is distinct information in gain-insensitive M/EEG field patterns arising from sparse subcortical and cortical source spaces. While simple physical laws suggest that electromagnetic field patterns can be related to the depth of single current sources [112–114], the degree to which this information allows disentangling/resolution of subcortical currents (by anatomic region) in a distributed source space across the brain has remained questionable. For example, explorations of distinctions in subcortical and cortical field patterns have found moderate to high correlations across structures [116]. However, these computations were limited to full anatomic structures, such as the full caudate or the full thalamus, which are typically not entirely activated at any given time. Our field comparison results, on the other hand, consider neurophysiologically relevant sparse subdivisions of brain structures, and reveal stark distinctions in their field patterns, with similar trends across both cortical and subcortical sources. Thus, our work suggests that information in field patterns can be utilized with sparse constraints to address the subcortical source modeling problem.

3.7.2 Applicability for Solutions to Subcortical Inverse Problem

Our observations are relevant to the natural sparsity in brain dynamics, and thus translate readily to practicable approaches for subcortical source estimation. Neurophysiologic processes often comprise relatively focal activity - involving only a select few brain regions at any given time, and evolving sparsely over space and time. For example, the somatotopic organization of the central nervous system is well known for functions such as sensation, language, and motor function, and specific regions of the brain are known to be involved in cognitive and emotional states such as memory, fear, reward and arousal [107, 110, 111]. As such, physiological sparsity in brain dynamics has

been increasingly recognized as a means to constrain the electromagnetic inverse problem. This has led to the emergence of techniques employing a number of sparsity-based priors and search approaches for cortical source estimation. These works use spatial sparsity priors, time-frequency sparsity constraints, or combinations therein, within weighted or penalized least squares, convex optimization, or pursuit-based search procedures [86, 190, 193–198] to solve the cortical inverse problem. Our field map comparison results suggest that these existing methods can now be adapted to offer a range of practical options for solving the subcortical inverse problem.

3.7.3 A Practical Algorithm for Subcortical Source Estimation

For a first exploration of feasibility, we adapted pursuit-based methods within a hierarchical search for sparse sources whose field patterns best match the data. Our results demonstrate that this hierarchical reduction approach affords the ability to track dynamics in specific thalamic and brainstem regions with good spatial resolution. Previous attempts to localize subcortical sources have used the naive minimum norm solution on the full source space and found poor spatial resolutions for deeper sources [119, 175–177]. By contrast, our algorithm offers two novelties. First, we start with the minimum norm solution and refine it using subspace pursuit concentrating estimates into regions whose field patterns best match the measurement. Second, we iteratively refine this set of possible sources across hierarchies of sources, allowing subcortical sources to enter the estimates. These novelties enable superior resolution, accuracy and performance than the have been previously reported [112–114, 117, 119, 121].

The resolution of our technique is determined by (a) the spatial separation required for high principal angles and (b) the relative current densities of the region of interest. First, sources that are physically close tend to have correlated field patterns. Thus, for mean principal angles $> 30°$, cortical sources need to be separated by $\sim 1.4\,\text{cm}$ and subcortical sources by $\sim 2\,\text{cm}$. Second, unlike the cortical mantle, subcortical volumes have non-homogenous current densities – and thus resolutions too vary across regions. For example, striatal regions have an order of magnitude greater current densities than thalamic sources [119], and thus can be resolved more finely, i.e., less volume for same current strength. Thus, a resolution on the range of $1.5 - 2.5\,\text{cm}$ is achievable depending on structure of interest, a similar order as resolution achievable for cortical sources [5, 13, 84–86]. Further, our method is practical as the hierarchical reduction of source spaces makes for a computationally efficient implementation, with typical runs for a 1 second data segment taking < 5 minutes across all stages. To the best of our knowledge, this is the first report of subcortical localization in a general distributed source model tailored to exploit differences in field patterns.

3.7.4 Implications for Neuroscience Studies

In summary, our results establish the feasibility of resolving neuronal currents within subcortical structures using non-invasive M/EEG recordings. As direct measures of electromagnetic activity within the brain are inaccessible with present experimental techniques, this framework opens

up unique opportunities to study regional dynamics across the brain at high temporal resolution. Example applications include studies of thalamocortical oscillations, which often exhibit high synchrony and generate large measurements [108, 109, 119, 199], studies of millisecond timescale functional connectivity or coherence across brain networks [40, 200–204], and dynamics underlying processes, such as arousal, emotion and memory, that critically involve subcortical regions. Further, our results suggest that using simultaneously recorded M/EEG may be better poised for subcortical source modeling than MEG or EEG alone. While EEG can measure higher amplitude contributions from deep sources than MEG can, the EEG field patterns are typically less distinguishable than MEG fields. Thus the two have complementary advantages for a subcortical inverse solution.

3.7.5 Future Directions

Future investigations will study improved ways to combine MEG and EEG data for this problem. Further enhancements in spatial resolution within the subcortical volume can also be anticipated with the development of methods to exploit functional or diffusion MRI based priors for reduction of source spaces. Expanded applications incorporating time frequency structure of the data and alternate ways to constrain sparsity in non-spatial bases can also be envisioned.

3.8 Appendix

3.8.1 Source Space Characteristics

Structures	**Type of Source Space**	**Total Size**	**Current Density**	**Number of Subdivisions**
Neocortex	Surface	$750\,\mathrm{cm}^2$	$0.25\,\mathrm{nAm/mm}^2$	642
Caudate	Volume	$3200\,\mathrm{mm}^3$	$0.25\,\mathrm{nAm/mm}^3$	27
Hippocampus	Surface	$4.1\,\mathrm{cm}^2$	$0.4\,\mathrm{nAm/mm}^2$	42
Inferior Colliculus	Volume	$390\,\mathrm{mm}^3$	$0.25\,\mathrm{nAm/mm}^3$	1
Lateral Geniculate	Volume	$120\,\mathrm{mm}^3$	$0.25\,\mathrm{nAm/mm}^3$	1
Medial Geniculate	Volume	$150\,\mathrm{mm}^3$	$0.25\,\mathrm{nAm/mm}^3$	1
Putamen	Volume	$4750\,\mathrm{mm}^3$	$0.25\,\mathrm{nAm/mm}^3$	36
Thalamus	Volume	$8960\,\mathrm{mm}^3$	$0.025\,\mathrm{nAm/mm}^3$	8

Table 3.1: Anatomical and Electrophysiological Properties of Source Space. Structure dimensions, based on the FreeSurfer template anatomy, are listed for left sided structures. Number of neocortical patches is listed for $\mathbb{C}^{(3)}$. Current densities, based on [119], were set symmetrically. When current densities were unavailable in literature, they were set based on neighboring regions. For surface sources, the number of patches is fixed by the icosahedron-based patch decomposition within the MNE software package [82, 181]. For volume sources, the number of subdivisions is derived as detailed in § 3.2.2.

3.8.2 Mutual Coherence Calculations

Mutual coherence thresholds μ are used in the subspace pursuit algorithm to enforce incoherence during search [86]. To estimate these thresholds, the maximum correlations between the modes of forward solutions from pairs of neighboring cortical patches were computed. These were first averaged within a neighborhood to obtain the average neighborhood maximum correlation, and then across neighborhoods to obtain the threshold μ.

In essence, these thresholds ensure that at each iterative subspace pursuit stage, correlation between the current support and new regions entering the solution is $< \mu$, preventing new regions from clustering around neighbors of the current support. Rather, it enforces that new regions entering the solution explain substantially different field patterns than the current support and its neighbors can. For a given subject, one threshold was set for each cortical patch decomposition $\mathbb{C}^{(1)}$, $\mathbb{C}^{(2)}$, and $\mathbb{C}^{(3)}$, and used for the successive hierarchies during cortical source space reduction. The final joint cortical and subcortical estimation stage used the threshold from the last cortical stage to allow any of the cortical sources in reduced source space $\mathbb{C}_{\text{sp}}$ to enter the final solution. Beyond this, the exact value of the threshold does not matter at the final stage because the sparse subcortical and cortical field patterns are implicitly nearly orthogonal. The mutual coherence values are summarized in Table 3.2.

Test Case	**Measurements Considered**	**Cortical Patch Decomposition**	**Maximum Mutual Coherence**	**Minimum Orthogonality Requirement (Degrees)**
Somatosensory Evoked Response	MEG	$\mathbb{C}^{(1)}$	0.75	41.4
		$\mathbb{C}^{(2)}$	0.83	33.9
		$\mathbb{C}^{(3)}$	0.88	28.4
Auditory Evoked Response	MEG and EEG	$\mathbb{C}^{(1)}$	0.76	40.5
		$\mathbb{C}^{(2)}$	0.86	30.7
		$\mathbb{C}^{(3)}$	0.89	27.1

Table 3.2: Mutual Coherence Thresholds for Subspace Pursuit. The thresholds were computed using whitened modes in each case, such that units were consistent across magnetometers, gradiometers and EEG. When MEG and EEG were considered jointly, the columns of the whitened EEG forward matrix and whitened MEG forward matrices were concatenated (across sensors) for each patch of interest. Forward solutions from neighboring patches are less correlated for coarser patches than for finer patches, thus the maximum coherence allowed increases with increasing fineness. This causes the search to enforce greater orthogonality in coarser source spaces allowing the algorithm to search widely before settling in on relevant regions. Finer source spaces allow more clustering, as the solutions are getting narrowed into the relevant regions.

3.8.3 Measurement Noise Covariance Estimates

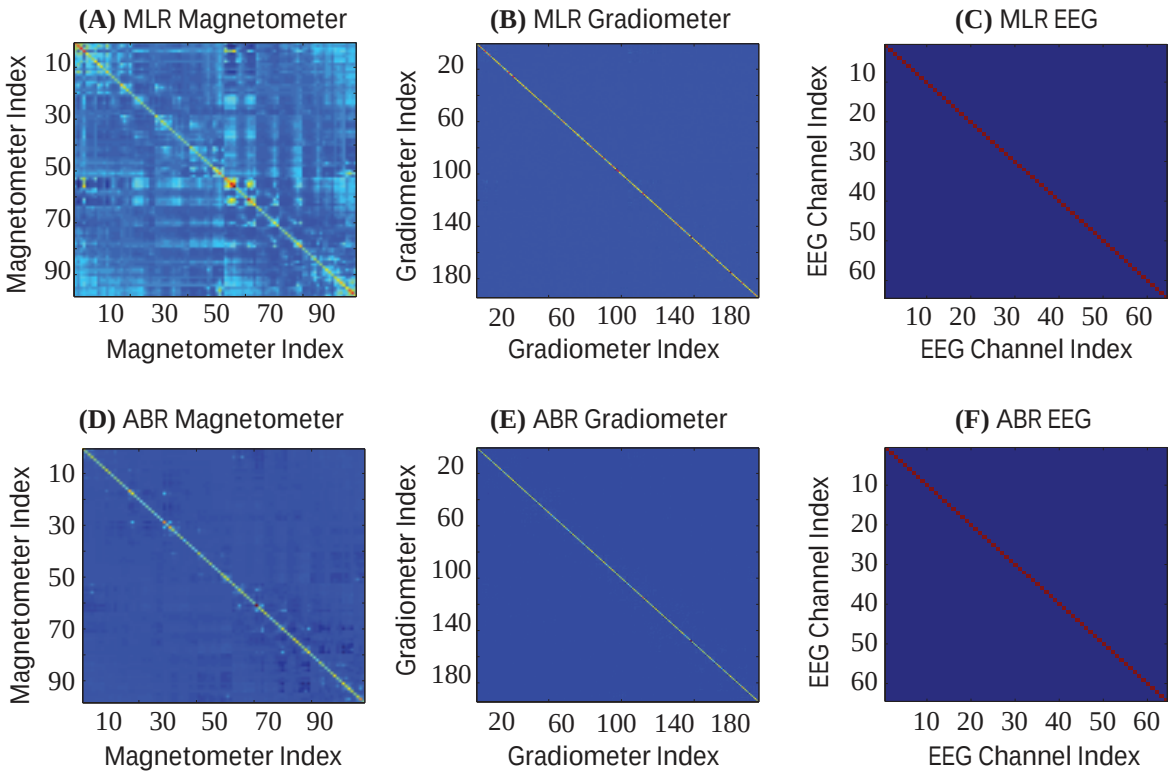

Figure 3-15: Example Estimates of Observation Noise Covariance. Regularized observation noise covariance matrices $\mathbf{Q}$ estimated with 2 minutes of eyes open baseline data. Data segments used for estimation were filtered to the frequency band under study, covariances were estimated within the MNE Software Package [82, 181], and then regularized by applying the SSP operator and diagonal loading (details in § 3.5.2). As reliable estimation of the EEG noise covariance matrix is difficult, the EEG noise covariance estimate was loaded heavily to reflect a mean diagonal value. In each case, the plotted quantity is $\sqrt{|Q_{j,k}|}$ where j and k denote sensor indices. *(A-C)* Covariances in the auditory middle latency response (MLR) frequency band $0-300\,\mathrm{Hz}$ for magnetometer, gradiometer and EEG channels with diagonal values $\sim 0.55\,\mathrm{fT}$, $\sim 45\,\mathrm{fT/m}$, and $\sim 4\,\mathrm{nV}$. *(D-F)* Covariances in the auditory brainstem respnose (ABR) frequency band $500-1625\,\mathrm{Hz}$ for magnetometer, gradiometer and EEG channels with diagonal values $\sim 0.71\,\mathrm{fT}$, $\sim 63\,\mathrm{fT/m}$, and $\sim 9\,\mathrm{nV}$. The magnetometer covariance has the most off-diagonal components, many of which are significantly reduced at the higher frequency range. For reference, these covariances were used for the auditory brainstem evoked analysis in § 3.5.2.

Chapter 4

Spatiotemporal Brain Dynamics Under Propofol General Anesthesia

In this chapter, we illustrate applications to studies of spatiotemporal dynamics underlying brain states of interest in cognitive and clinical neuroscience. Specifically, we study brain dynamics underlying loss of consciousness induced by propofol general anesthesia. Anesthetic drugs like propofol have established mechanisms of action at the molecular level. Further, reliable surface EEG signatures demarcating propofol-induced loss of consciousness have been characterized. However, regional and systems-level brain dynamics associated with these molecular actions and EEG signatures are poorly understood. Spatiotemporal imaging offers a means to empirically connect regional and systems-level brain dynamics with EEG signatures and human behavior. Here, we analyze simultaneous EEG-fMRI data acquired during induction of propofol anesthesia to relate drug-induced oscillatory EEG signatures with auditory fMRI responses. We begin by applying the algorithm developed in Chapter 2 to remove ballistocardiogram artifacts from EEG recorded in the scanner. Next, we perform detailed time-frequency analyses on the cleaned EEG, to derive oscillatory EEG signatures that characterize propofol-induced unconsciousness in the scanner. We then proceed to associate these EEG signatures with auditory fMRI responses. We find that onset of slow and frontal alpha oscillations at clinical loss of consciousness corresponds to suppression of the auditory response pathway. Further, we observed that coupling between slow wave peaks and alpha oscillation amplitude during profound unconsciousness, well after clinical loss of consciousness, is associated with a state of intense cortical inhibition and subcortical reactivation. Our results enhance understanding of anesthesia-induced changes in brain function, and provide neurophysiologic correlates to better interpret clinical EEG signatures demarcating states of consciousness under anesthesia.

The propofol EEG-fMRI recordings studied in this work were acquired by Patrick Purdon, Eric Pierce, Giorgio Bonmassar, John Walsh, Grace Harrell, Jean Kwo, Daniel Deschler, Margaret Barlow, Rebecca Merhar, Camilo Lamus, Catherine Mullaly, Mary Sullivan, Sharon Maginnis, Debra Skoniecki, Helen-Anne Higgins and Emery Brown at the Massachusetts General Hospital (MGH) [63]. The fMRI analyses were performed by Patrick Purdon at the MGH Martinos Center for Biomedical Imaging.

4.1 Introduction

General anesthetic drugs are used in millions of medical and surgical procedures to induce loss of responsiveness, awareness and consciousness [24]. However, the mechanisms by which these dramatic changes in brain function come about remain incompletely understood. Experimental studies of these mechanisms have been limited by challenges in imaging spatiotemporal dynamics across extensive brain regions (Chapter 1). Further, monitoring brain state under general anesthesia remains to be adopted as routine clinical practice [150]. Clinically observable markers of depth of anesthesia, on the EEG for instance, do not have established associations with underlying regional and network dynamics, or with specific disruptions in cognitive processing pathways. Thus, there is a need to characterize how anesthetic drugs reconfigure brain dynamics, and associate these changes in brain dynamics with clinically observable EEG signatures.

Propofol (2,6-di-isopropylphenol) is one of the most commonly used general anesthetic drugs. It is administered intravenously for induction and maintenance of general anesthesia – as it induces hypnosis and amnesia [205]. It is also used for sedation in procedural and intensive care settings as it has the desirable characteristics of rapid onset and recovery [24, 205]. At the molecular level, propofol potentiates the brain's primary inhibitory neurotransmitter, gamma-aminobutyric acid (GABA) [24, 206], and is known to act on GABA-A receptors widely distributed in cerebral cortex, thalamus, brainstem, hippocampus, hypothalamus, basal ganglia, and cerebellum [207–209]. The expansive spread of receptors suggests a multitude of possible circuit mechanisms by which propofol can alter brain dynamics and cognitive state.

Macro-level changes in brain dynamics induced by propofol have been characterized by a number of EEG studies [191, 210–212]. Specifically, propofol-induced unconsciousness has been linked to a widespread shift in EEG power toward lower frequencies [150, 213, 214]; increases in frontal EEG power [213, 215–217] with particularly high spatial coherence (or synchrony) in the alpha frequency range ($8 - 12\,\text{Hz}$) [150, 213, 218]; emergence of a slow ($< 1\,\text{Hz}$) oscillation that is asynchronous across cortex [150, 214]; and onset of burst suppression at the deepest doses [219, 220]. Stepwise variations in propofol dose alongside auditory response tasks have been employed to identify EEG signatures that punctuate loss of responsiveness and awareness, as summarized in Fig. 4-1. Specifically, the loss of behavioral responsiveness has been associated with increase in slow-band and frontal alpha-band power [91, 150, 221]. Further, behaviorally non-apparent transitions in consciousness have been associated with two distinct cross-frequency coupling signatures [91, 150, 214]. First, coupling between troughs of the slow oscillation and amplitude of the alpha oscillation predicts transitions into unconsciousness. Second, coupling between peaks of the slow oscillation and amplitude of the alpha oscillation marks profoundly unconscious states where awareness is believed to be unlikely [150]. Overall, anteriorized hypersynchronous alpha oscillations, widespread asynchronous slow oscillations, and phase-amplitude coupling between these oscillations constitute known EEG signatures demarcating propofol-induced changes in conscious cognitive processing.

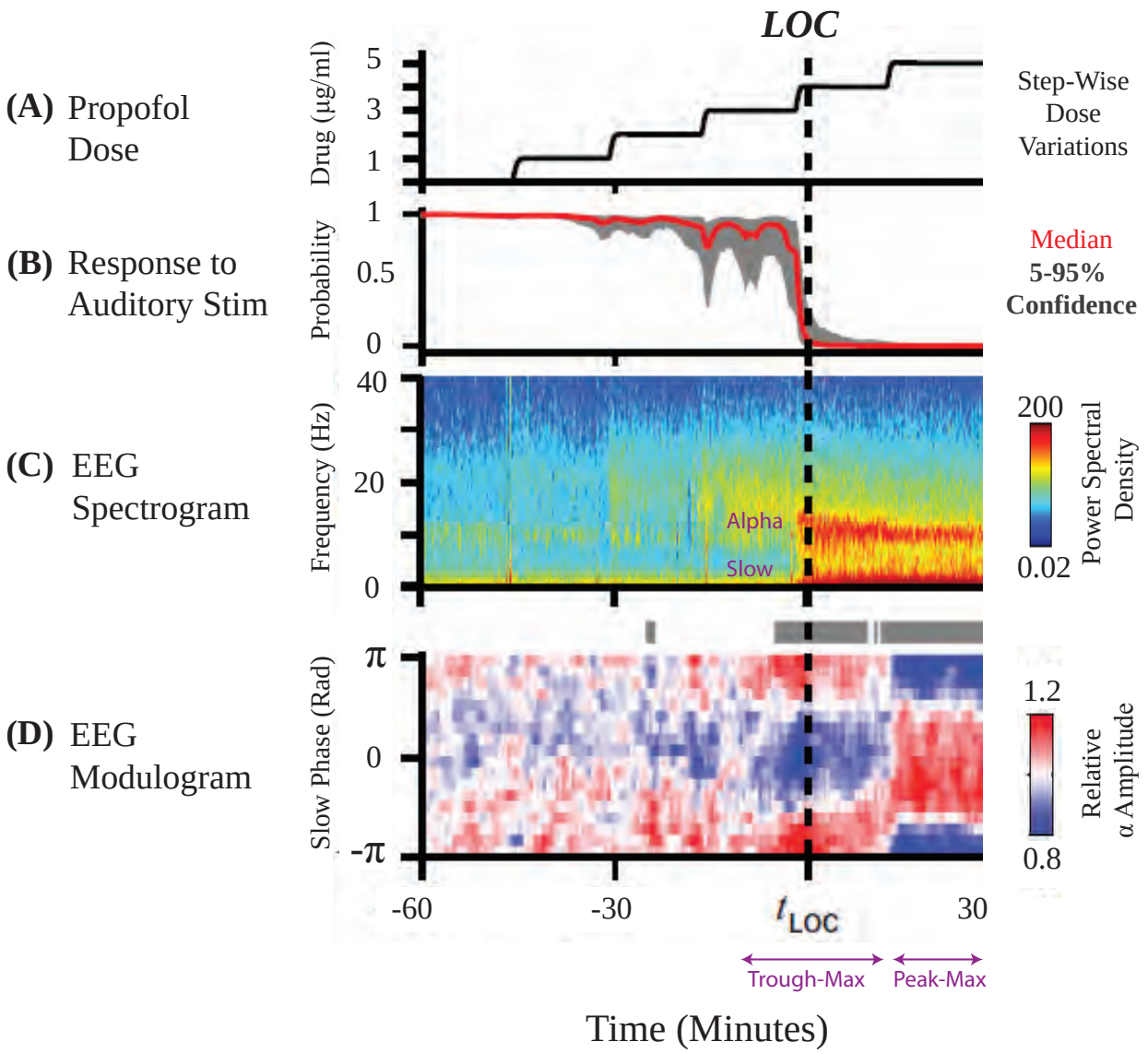

Figure 4-1: Known EEG Signatures Marking Propofol-Induced Behavioral Changes. Summary results from EEG studies on healthy volunteers during propofol anesthesia, adapted from [91, 150]. *(A)* Propofol induction with gradual dosing in step-wise increments over time. *(B)* Probability of response to a behavioral task. Loss of behavioral response is marked as clinical loss of consciousness (LOC). *(C)* EEG spectrogram on representative frontal channel shows a rapid increase in oscillatory power in the slow and alpha frequency ranges at LOC. The oscillations are sustained with increasing drug dose after LOC. *(D)* EEG modulogram quantifying the phase-amplitude coupling between the slow and alpha oscillations (waxing and waning of the power in the alpha frequency band at specific phases of the slow rhythm). Profound levels of unconsciousness are marked by a peak-max modulation pattern.

The highly structured nature of EEG oscillations marking propofol-induced behavioral and cognitive states suggests that these states are associated with large-scale systems-level dynamical changes, and not sporadic molecular actions. But, how propofol reconfigures dynamics in specific brain regions or networks in association with these known EEG and behavioral changes remains unresolved.

Cortical dynamics associated with these oscillatory rhythms have been studied with electrocorticography and single unit recordings in clinical patient populations [214, 222], as well as with EEG source imaging in healthy volunteers [91, 221]. However, dynamics involving subcortical structures (brainstem and thalamus) and their interactions with the cortex, known to play an important role in propofol unconsciousness [206, 223], remain difficult to resolve. PET studies have been performed to characterize changes in brainstem, thalamus and cortex during general anesthesia [224–227]. Notably, these studies have demonstrated suppression of midbrain and thalamic activity during unconsciousness, and revealed activation in anterior brain arousal networks during emergence. However, the low spatiotemporal resolution of PET limits applicability for characterization of regional dynamics tied to loss of consciousness [214]. As fMRI has better spatiotemporal resolution, it has been used to study propofol-induced connectivity changes in corticocortical (particularly frontoparietal and default mode) networks, as well as in thalamocortical networks [53, 228–231]. While these studies have identified suppression of higher-order cortical networks and nonspecific thalamic nuclei as correlates of unconsciousness, they do not systematically address questions relating to changes in stimulus processing during and after loss of consciousness. Further, standalone fMRI lacks temporal resolution for association with in vivo electrophysiologic correlates or clinically observable EEG signatures.

As a consequence, while the EEG signatures are easily observed in the clinical setting, their interpretation is limited to connections with observable behaviors. How these EEG signatures relate to non-apparent functional brain states and associated dynamical changes within specific brain pathways or networks remains to be characterized. Specifically, associations between propofol EEG oscillations and suppression of specific stages in the hierarchy of sensory processing remain unresolved. For example, how propofol affects thalamic relay of sensory information is a matter of debate. Several reports have shown reduction in thalamic activity during propofol-induced unconsciousness [206, 224–226, 232]. However, whether this reduction is a cause of suppressed routing of sensory information to cortex, or an effect of propofol-induced changes in cortical activity remains unclear [221, 233]. Similar questions prevail with regards to how propofol EEG oscillations affect memory forming capacity and other innate cognitive functions. Thus, there is a need to empirically characterize how clinical EEG signatures relate to changes in brain pathways for sensory processing, memory formation and other cognitive functions under propofol anesthesia.

Spatiotemporal imaging offers a means to empirically connect brain dynamics with electrophysiologic recordings and human behavior. Specifically, simultaneous EEG- fMRI can relate fast timescale dynamics and clinically observable signatures on the EEG with their underlying fMRI-based regional or network changes. While clinical and technical challenges in acquiring joint EEG-fMRI data during induction of deep propofol anesthesia have been overcome [63], significant technical challenges have persisted in analyzing and interpreting EEG acquired in the MRI scanner, alongside fMRI-based regional dynamics.

The advances in simultaneous EEG-fMRI artifact removal from Chapter 2 provide a means to overcome these challenges. In this chapter, we analyze simultaneous EEG-fMRI data acquired during propofol anesthesia [63] to relate drug-induced oscillatory EEG signatures with fMRI-based regional dynamics. First, we apply the harmonic regression techniques from Chapter 2 to remove ballistocardiogram artifacts from EEG recorded in the MRI scanner during propofol anesthesia. Second, we perform spectral and cross-frequency coupling analyses on the cleaned EEG estimates to derive oscillatory EEG signatures that characterize propofol-induced unconsciousness in the MRI scanner. Third, we associate these EEG signatures with fMRI-based sensory responses at progressive levels of propofol-induced unconsciousness. We conclude by outlining opportunities for future EEG-fMRI studies aiming to characterize regional dynamics underlying propofol-induced EEG oscillations.

4.2 Methods

4.2.1 EEG-fMRI Data Acquisition

The data used in this chapter were obtained on volunteer subjects undergoing interleaved EEG-fMRI during the induction and maintenance of propofol general anesthesia [63]. This study was approved by the Partners Human Research Committee at Massachusetts General Hospital (protocol 1999-P-010748), and all subjects provided written informed consent. The study enrolled 7 volunteers (50-60 y/o) who had preexisting tracheal stomas allowing awake intubation prior to positioning in the scanner. Subjects were recruited with support from the International Association of Laryngectomies, and screened to ensure (a) an ASA score $>$ II, (b) no neurological or psychiatric abnormalities and (c) no MRI contraindications.

Fig. 4-2 illustrates the study paradigm. Propofol was administered intravenously with a computer controlled infusion pump that progressively increased targeted effect-site concentration in step-wise increments of 1 ug/mL. Each dose level was maintained for 15 minutes. During the infusion and maintenance period, a 500 millisecond tone was presented once every minute. The tone was randomized to low or high pitch (220 Hz vs. 440 Hz) and subjects were instructed to identify pitch with a button press response. Reaction time and accuracy of the response were recorded. EEG was continuously recorded at 950 Hz with bipolar electrodes at 24 locations across the head, configured per the 10/20 layout. Functional MRI was acquired using a Trio^{TM} 3 T scanner (Siemens, Erlangen, Germany). To minimize interference between the auditory stimuli under study and acoustic scanner noise, a sparse sampling paradigm was used. Each fMRI volume (4-mm slice thickness, 1-mm skip, 3.1$\times$3.1-mm in-plane resolution, 64 $\times$ 64 matrix, TE = 30, 90-degree flip angle) was acquired in 1 second, followed by $8-10$ second silent periods. Throughout the study, end-tidal carbon dioxide was maintained within 5% of baseline to ensure physiologic stability and image quality. Physiologic variables were monitored as is standard during clinical administration of general anesthesia. All anesthesia and monitoring equipment were MRI compatible. At the end of the study, subjects were brought outside the MRI scanner to emerge from general anesthesia.

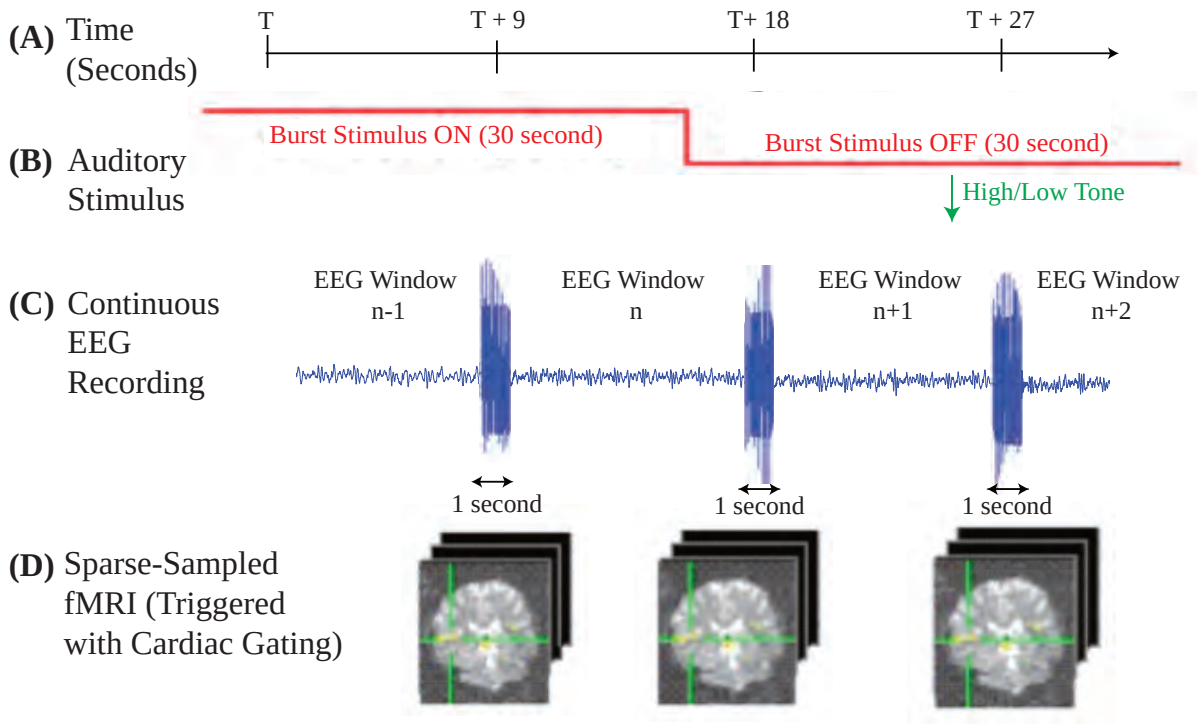

Figure 4-2: Acquisition of EEG-fMRI Data During Propofol Anesthesia. Subjects lay in the scanner with eyes closed as propofol anesthesia was induced and maintained with gradual step-wise dose increments over a period of 2 hours. The study design is illustrated for a short segment of time (*A*). *(B)* A 12.5 millisecond auditory burst stimulus repeating at 40 Hz was presented every 30 seconds. When the burst stimulus was off, a tone randomized to high or low pitch associated with a behavioral task was presented. *(C)* EEG was recorded continuously during the induction-behavioral paradigm. *(D)* Each fMRI volume covered coronal slices spanning brainstem, midbrain and auditory cortex, and was acquired in 1 second. The accompanying gradient artifacts corrupting the EEG are seen in *(C)*. Each volume acquisition was followed by a 7.5 second delay. Then, cardiac gating was enabled to trigger the subsequent volume acquisition with minimal motion artifacts. Effective TR was $\sim$ 9 seconds. Figure is adapted from [63] which reports acquisition of the data that we analyze in this chapter.

4.2.2 Behavioral Analysis

Clinical loss of consciousness (LOC) was defined as the time point at which the subject stopped responding to the auditory tone stimuli. The subject's behavioral state was defined relative to the propofol dose level at which LOC occurs. The levels prior to LOC constitute the conscious baseline state, where the subject had eyes closed, but was either awake or behaviorally responsive during induction. For each subject, two dose levels prior to LOC was defined as LOC-2 and the dose level just prior to LOC was defined as LOC-1. The levels following LOC constitute unconscious states. The dose level just after LOC was defined as LOC+1, and constitutes the transition into unconsciousness. Two dose levels after LOC was defined as LOC+2. LOC+2 and beyond constitute the profoundly unconscious state.

4.2.3 EEG Spectral Analysis

To relate propofol-induced oscillatory EEG signatures with fMRI-based regional dynamics, we analyzed the EEG data recorded alongside fMRI at 3 T. All EEG analyses were performed with the native bipolar reference used in the data collection. Each channel of EEG data was analyzed independently of other channels. Two subjects were excluded from the EEG analysis: one of whom had large gross motion artifacts corrupting the EEG data, while the other was not induced to levels of unconsciousness where slow oscillations developed. In descriptions of EEG analyses and results, we use the 'slow' band for $0.1 - 1$ Hz, and the 'alpha' band for $8 - 14$ Hz.

Spectrograms were computed using multitaper methods implemented within the Chronux toolbox (http://chronux.org/) [234]. EEG spectrograms used 3 second long non-overlapping data segments between fMRI volume scans. Consecutive data windows were separated by the effective TR of ~ 9 seconds (as shown in Fig. 4-2). The number of tapers was set to 4 and the time-bandwidth product was set to 2.5 for a spectral resolution of 1.6 Hz. Power in the slow (0.1-1 Hz) and alpha (8-14 Hz) bands of interest were computed as total power spectral density, numerically integrated across frequencies within the respective bands. For band power plots, time points when $> 75\%$ of the power spectral densities at frequencies below 25 Hz had outlying values $> 90^{\text{th}}$ percentile of all power spectral density values were considered artifactual and excluded. Then, visual inspection of raw time series was used to exclude any other periods when gross motion artifacts were present.

4.2.4 Removal of Ballistocardiogram Artifacts

As static field ballistocardiogram artifacts pose challenges in discerning propofol-induced slow and alpha band oscillatory dynamics in raw EEG collected at 3T, we applied the harmonic regression techniques developed in Chapter 2 to remove artifacts, and recover true brain generated EEG rhythms. Artifact removal was performed on 3 second long moving windows at all levels. Model selection for the harmonic and autoregressive orders was performed on 5 representative data segments across 2 channels for each of the 5 propofol dose-levels – totaling 50 data segments per subject. For each subject, one set of model orders was chosen across all levels. Typical model orders for our subjects were $R = 20$ harmonics and $P = 4 - 8$ autoregressive terms. Goodness of fit was assessed alongside model selection (§ 2.4.2.2) to ensure the chosen model sufficiently explains the data. For a chosen model order set $[R, P]$, sensitivity to model orders was assessed by comparing (a) estimated fundamental frequencies and (b) goodness of fit results for harmonic orders ranging $[R - 3, R + 3]$, and autoregressive orders ranging $[P - 3, P + 3]$. Artifact removal results were relatively insensitive to model choice within this ballpark range. To facilitate efficient analysis of large volumes of data, EEG data were downsampled by 8X, dropping effective sampling rate of 119 Hz, for the 1-D optimization of likelihood across frequency (Eqn. 2.16). Fully sampled EEG data were used to perform the other optimization steps. This made computations faster without compromising convergence. Data segments analyzed and spectrogram settings were identical for raw and cleaned EEG.

4.2.5 Phase-Amplitude Coupling Analysis

To analyze the coupling between slow oscillation phase and alpha oscillation amplitude, standard phase-amplitude coupling quantifications [91, 98, 235] were adapted to the short discontiguous EEG segments available for our analysis, as detailed in the three steps below.

First, the slow oscillation phase and alpha oscillation amplitudes were computed. The cleaned EEG data segments (each of length 3 seconds) were bandpass filtered to extract slow-band ($0.1 - 1$ Hz) and alpha-band ($8 - 14$ Hz) signals. Symmetric finite impulse response filters designed with sharp transition bands to match the ideal response per a least squares criterion (MATLAB firls) were used. Then, a Hilbert transform was applied on each segment to compute the instantaneous slow oscillation phase $\phi(t)$ and the alpha oscillation amplitude $A(t)$. At each instant, the phase value $\phi(t)$ was assigned to one of $N = 100$ equally spaced phase bins dividing $[-\pi, \pi]$. The phase bin corresponding to $\phi(t)$ is denoted as $\phi_{\mathbf{b}}(t)$.

Second, a modulogram or a normalized distribution of amplitudes over phase bins was obtained as detailed in [235]. For a given slow phase $\phi(t)$, the alpha amplitude $A(t)$ has a probability density specified by $p(A;\, \phi)$, and the phase-amplitude coupling relation is $A(\phi,\, t) = E_p(A(t)|\phi)$ where E_p denotes ensemble averaging with respect to the density $p(A;\, \phi)$. When there is a coupling between the alpha amplitude $A(t)$ and the periodic slow phase $\phi(t)$, $A(\phi,\, t)$ is periodic. Thus, assuming local stationarity, the function $A(\phi,\, t)$ can be represented in the Fourier basis as:

$$A(\phi,\, t) = \mu + \sum_k a_k \sin\left(k\phi(t)\right) + b_k \cos\left(k\phi(t)\right).$$

$A(\phi,\, t)$ was estimated using a first order Fourier expansion $k = 1$, and a simple linear regression fitting the amplitude time series $A(t)$ to the phase-based regressors $[\sin\left(\phi(t)\right),\, \cos\left(\phi(t)\right)]$. For this estimation, the amplitude and phase time series were pooled across 40 data segments, each of length 3 seconds, so as to consider 2 minutes of data around each time point t. Then, at each time point, the modulogram for a given phase bin $\phi_{\mathbf{b}}$ was computed as the average amplitude for all samples $A(\phi_{\mathbf{b}})$ within that phase bin $\phi_{\mathbf{b}}$, normalized by the average amplitude across all phase bins. Thus, modulogram $M(t,\, \phi_{\mathbf{b}})$ provides a normalized distribution of alpha amplitudes over slow phase bins.

Third, the strength of the phase modulation was assessed by computing a modulation index $MI(t)$, which quantifies (in bits) the Kullback-Liebler divergence of the phase amplitude modulogram from the uniform distribution. Based on [98]:

$$MI(t) = \sum_{n=1}^{N} M(t,\, \phi_{\mathbf{b},\mathbf{n}}) \log_2 \frac{M(t,\, \phi_{\mathbf{b},\mathbf{n}})}{1/N}.$$

To assess statistical significance of the modulogram, a permutation test was employed. Amplitude segments were randomly time-shifted with respect to the phase segment, and a permuted modu-

lation index $MI_{\text{perm}}(t)$ was computed. 200 permuted modulation indices were computed for each phase segment. The empirical modulation index $MI(t)$ was deemed significant if it exceeded 95% of the permuted modulation indices $MI_{\text{perm}}(t)$.

4.2.6 Functional MRI Analysis

The fMRI data were analyzed previously [Purdon, 2009, unpublished data]. The fMRI time series were registered and motion corrected using standard procedures, and then statistically analyzed to characterize brain responses to the auditory tone stimuli. A general linear model framework implemented within the FSL software package (Analysis Group, FMRIB, Oxford, UK) was employed. The regressors comprised a convolution of a binary stimulus indicator with the hemodynamic response function. The contrasts of the regression parameters were characterized using a Markov-chain Monte Carlo-based mixed effects analysis [236]. This approach accounts for the fixed effects variance within an individual subject and the random effects variance for group analyses. For group analyses, the fMRI data were registered across subjects relative to the LOC point. Group fMRI analyses were conducted for each of the behaviorally-defined levels (LOC-2, LOC-1, LOC+1, and LOC+2). The resulting group activation map comprises voxel-wise partial F-statistics corrected for multiple comparisons, and is thresholded to ensure that group-level probability of false positives < 0.05. The threshold is uniform across all dose levels. As propofol does not significantly affect cerebrovascular coupling [237, 238], the fMRI results largely represent drug-induced changes.

4.3 Results

To relate EEG signatures and fMRI regional dynamics as a function of behavioral state, the propofol dose, behavioral response, EEG results, and fMRI results were all aligned in time and plotted through the course of the study. The time at which the subject stops responding is denoted as clinical loss of consciousness (LOC). In all plots, the drug, dose, behavioral response and EEG results are plotted for individual subjects. All EEG results are shown for a single representative frontal or occipital channel, as indicated on plots. The fMRI responses to auditory stimuli are presented as statistical maps obtained from the group analysis comprising 7 subjects.

Fig. 4-3 shows raw frontal and occipital EEG spectrograms alongside the behavioral and fMRI results. Fig. 4-3*A-B* allow behavioral categorization of the state of consciousness as it evolves through the study. Fig. 4-3*C-D* show the spectrograms computed using raw EEG recordings in the scanner. As the subject loses consciousness, it is possible to see reduced gross motion artifacts in these raw spectrograms, but it is hard to discern specific oscillatory EEG signatures that manifest with changes in dose and behavior, and relate them to fMRI-based regional dynamics in Fig. 4-3*E*. Thus, we removed ballistocardiogram artifacts from the raw EEG. Fig. 4-4 shows the revised plot. Fig. 4-4*C-D* shows the frontal and occipital EEG spectrograms after BCG removal (same subject and channels as in Fig. 4-3).

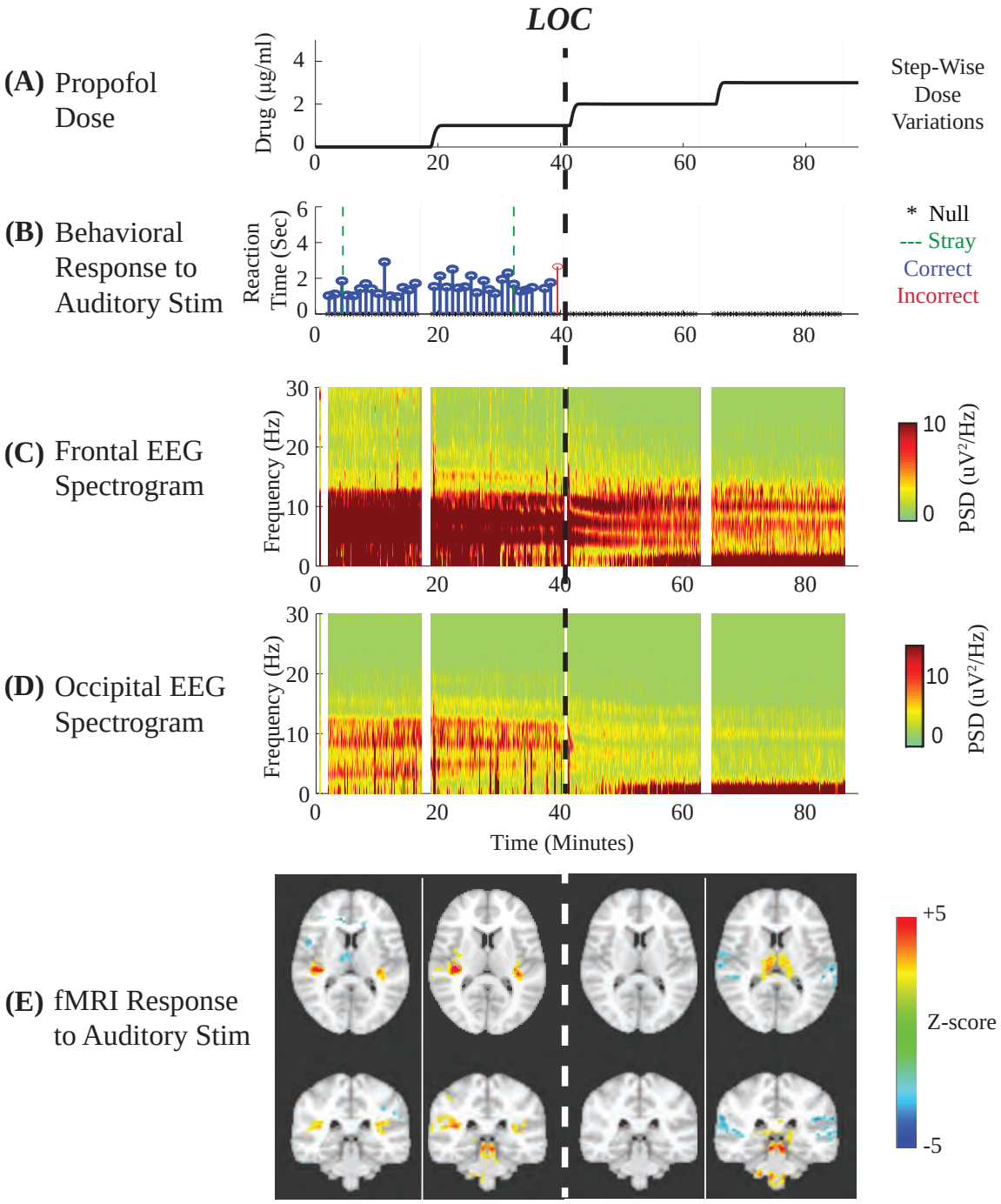

Figure 4-3: Raw EEG Spectrograms Alongside fMRI-based Regional Dynamics During Induction of Propofol Anesthesia. (*A*) Propofol induction with gradual dosing in step-wise increments over time. (*B*) Reaction time for response to the auditory tone stimulus plotted over time as the study progresses, color-coded based on presence and accuracy of response. (*C-D*) Raw EEG spectrograms through the study on a frontal channel (F4 →Fp2) and occipital channel (O1 →P7). Gaps between levels correspond to equilibration period between dose increments. (*E*) Group fMRI responses to the auditory stimuli. (*D-E*) Spectral changes associated with LOC and fMRI-based regional dynamics are hard to discern due to static field ballistocardiogram artifacts.

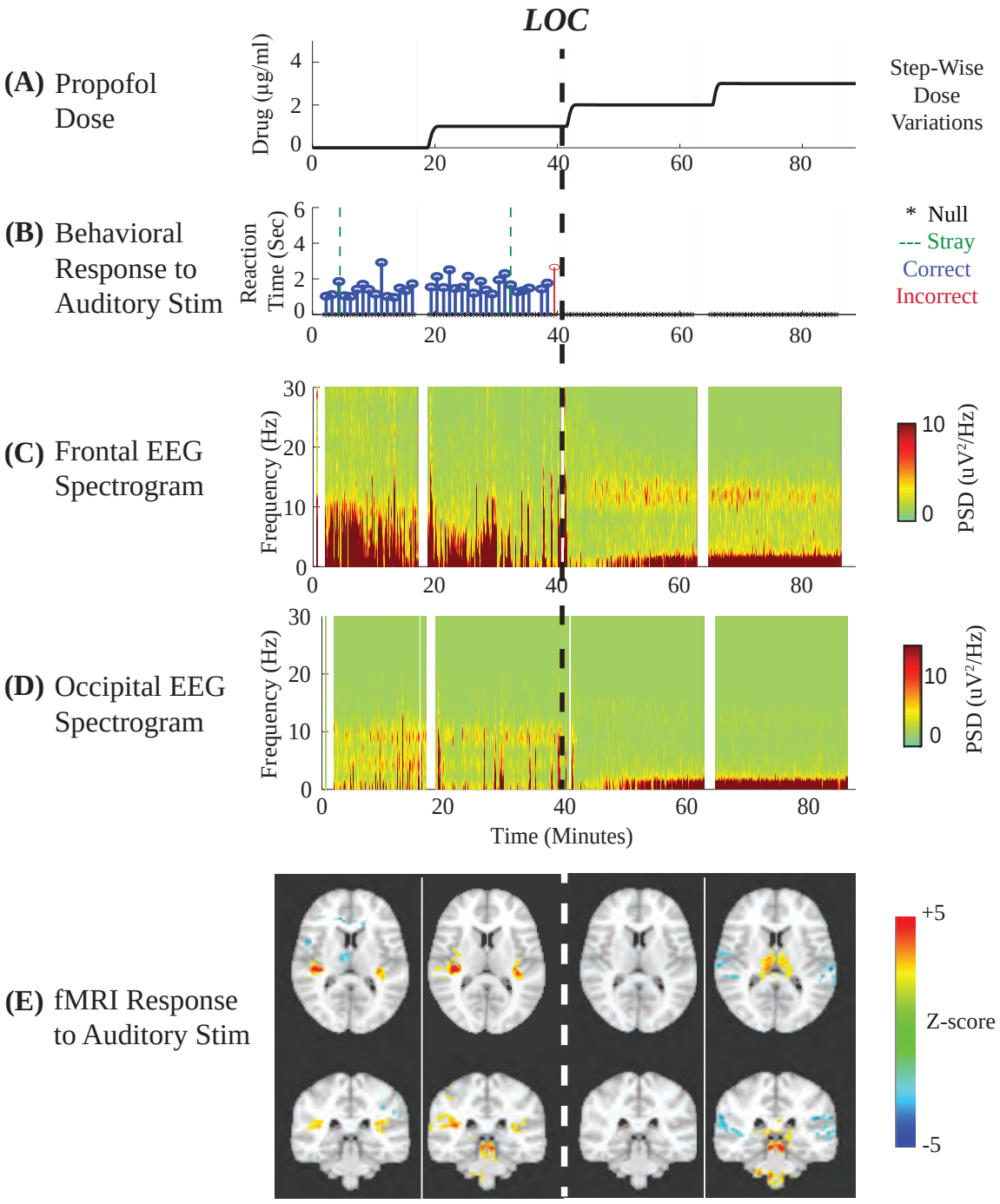

Figure 4-4: Clean EEG Spectrograms Alongside fMRI-based Regional Dynamics During Induction of Propofol Anesthesia. (*A-B*) Progressive propofol dosing and associated loss of behavioral response to auditory tone stimuli through the study. (*C*) Frontal EEG spectrogram (F4 →Fp2) shows motion artifacts in conscious baseline state. Slow and alpha oscillations set in at LOC and are sustained through deeper levels. (D) Occipital EEG spectrogram (O1 →P7) shows alpha oscillation and some motion artifacts in conscious baseline state. Alpha oscillations disappear upon LOC, while slow oscillations set in at LOC and are sustained through deeper levels. (*E*) Group fMRI responses to the auditory stimuli. (*D*-E) Baseline occipital EEG alpha oscillation is associated with intact fMRI responses in the auditory pathway. EEG spectral changes at LOC mark reduced fMRI responses in the auditory pathway.

Now, the propofol-induced oscillatory signatures are evident, just as seen in studies outside the scanner (Fig. 4-1), and can be associated with the fMRI-based regional dynamics (Fig. 4-4*E*) as a function of drug dose and behavioral state (Fig. 4-4*A*-*B*). Our observations associating the EEG and fMRI dynamics are grouped into (a) conscious baseline state (levels prior to LOC), (b) transition into unconsciousness (LOC+1, i.e., the dose immediately following LOC), and (c) the state of profound unconsciousness (LOC+2 and beyond).

4.3.1 EEG Spectral Changes Punctuate fMRI-based Regional Dynamics at Loss of Consciousness

Fig. 4-4 presents the behavioral state (Fig. 4-4*A*-*B*), EEG spectral dynamics (Fig. 4-4*C*-*D*) and fMRI-based regional dynamics (Fig. 4-4*E*).

In the conscious baseline state, when subjects are behaviorally responsive to auditory stimuli, the frontal EEG spectrogram does not contain significant oscillatory dynamics. Rather, gross motion artifacts are visible. During this period, the occipital EEG spectrogram shows alpha-range oscillations, which correspond to the eyes closed state. For these baseline levels, the fMRI analysis reveals activation in the primary auditory pathway (inferior colliculus, medial geniculate thalamus, and primary auditory cortex). Thus, the occipital alpha power is associated with active responsiveness in non-visual primary sensory pathways.

During the transition into unconsciousness (marked by LOC), the occipital and frontal spectrograms show EEG oscillatory changes with respect to the baseline state. There is a decrease in occipital alpha power and an increase in frontal alpha power. Further, in both frontal and occipital channels, we see the onset of a high amplitude slow oscillation. During the transition into unconsciousness (marked by LOC), the fMRI results show a suppression of activity in the primary auditory pathway. Thus, the emergence of the frontal alpha oscillation and widespread slow oscillations mark a suppression of responses in the primary sensory pathways.

During the unconscious states following LOC, the EEG spectrograms maintain the oscillatory features that set in at LOC, while the fMRI responses evolve as the level of anesthesia deepens from LOC+1 to LOC+2. Specifically, in the profoundly unconscious state (LOC+2), the EEG spectrograms continue to show sustained frontal alpha oscillation and widespread slow oscillations. However, as the level of anesthesia deepens to the profoundly unconscious state (LOC + 2), the fMRI results show (a) increased suppression of cortical responses (negative BOLD, particularly in secondary cortices) and, unexpectedly, (b) an increased activation of the subcortical regions (thalamus, medial geniculate nucleus and inferior colliculus) in response to stimuli. To better assess if any quantifiable EEG power changes occur between the LOC+1 and LOC+2 states, we plotted the band powers across subjects.

4.3.2 EEG Band Powers are Sustained even as fMRI Responses Change During Unconsciousness

Fig. 4-5 and Fig. 4-6 show alpha and slow band powers respectively, in relation to the behavioral states for each subject. Each panel of these plots shows the data for one subject. For example, Panel 1 in both Fig. 4-5 and Fig. 4-6 shows band power plots for Subject 1. As Fig. 4-4*E* shows that fMRI responses change during unconsciousness (from LOC+1 to LOC+2), we focus on EEG power changes through these graded levels of unconsciousness.

Fig. 4-5 shows the frontal alpha-band power in relation to the behavioral state. We see that frontal alpha-band power typically begins to rise at LOC (black dashed line), and sustains through LOC+2 (to right of gray solid line). In subjects 1 and 3, we note that the alpha-band powers start to level off or drop at LOC+2 (marked by gray solid line). However, this decrease in alpha-band power does not consistently mark the transition to profound unconsciousness (from LOC+ 1 to LOC+2, solid gray lines) across subjects.

Fig. 4-6 shows the occipital slow-band power in relation to the behavioral state. While channels across the head show large slow oscillations, the power trends are presented for occipital channels as frontal channels typically have low frequency eye movement artifacts in baseline states. In 3 out of 5 subjects, we see that slow-band power begins to rise at LOC (black dashed line), and sustains through LOC+2 (to right of gray solid line). In subjects 4 and 5, the slow-band power seems to rise only at LOC+2. Comparing these results to those outside the scanner under similar conditions [150], we note that this may be a consequence of the behavioral task used to assess LOC. Specifically, we note that if LOC were to be assessed with a more potent stimulus, such as a word identification task, the start of large amplitude slow oscillations might have aligned with the LOC point [150]. Further, it appears that the overall slow-band power does not rise significantly within the LOC+2 state (or beyond). However, the slow-band power does continue to fluctuate every few seconds through the LOC+2 state. In addition, a saturation of the slow-band power does not consistently mark the transition to profound unconsciousness (from LOC+ 1 to LOC+2, solid gray lines) across subjects. Thus, neither rise or saturation of slow-band power consistently marks the onset of the profoundly unconscious state.

The consistent general trend across the band power plots, is similar to that seen in spectrograms of Fig. 4-4 - namely that frontal alpha and widespread slow oscillations develop at LOC, and sustain through more profound states of unconsciousness. While these band-powers remain to be collated across subjects and channels for group summaries, we note that these trends are qualitatively similar to those seen outside the scanner [150, 214]. Further, the power changes with increasingly profound levels of unconsciousness (beyond LOC) are subtle and depend on the subject, the stimulus paradigm under study, and the recording context. Therefore, to better understand the profoundly unconscious state, we examined the coupling between slow and alpha oscillations for signatures demarcating the fMRI changes occurring at deeper levels.

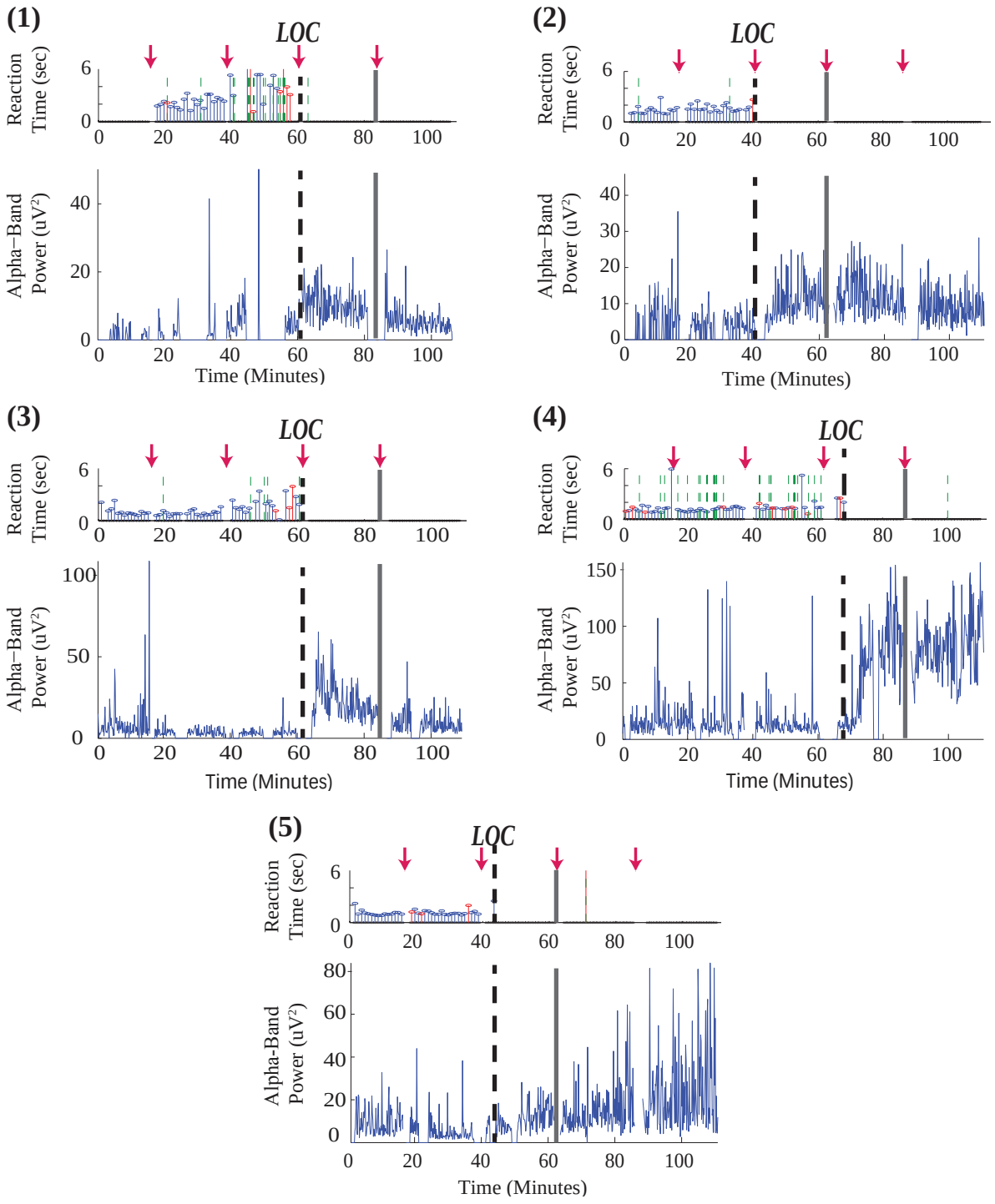

Figure 4-5: Frontal Alpha-Band Power Across Behavioral States for 5 Subjects. Each panel corresponds to one subject. Panel numbering corresponds to subject numbers used in text. In a given panel, the red arrows denote time points corresponding to drug dose increments and top subplots show behavioral responses. The dashed black lines mark LOC, defining transition into unconsciousness (LOC+1). The solid gray lines mark transition into profound unconsciousness (LOC+2 and beyond). The bottom subplots show evolution of the band powers ($8-14\,\mathrm{Hz}$) through behavioral states. We note that Fig. 4-3-Fig. 4-4 show spectrograms from Subject 2.

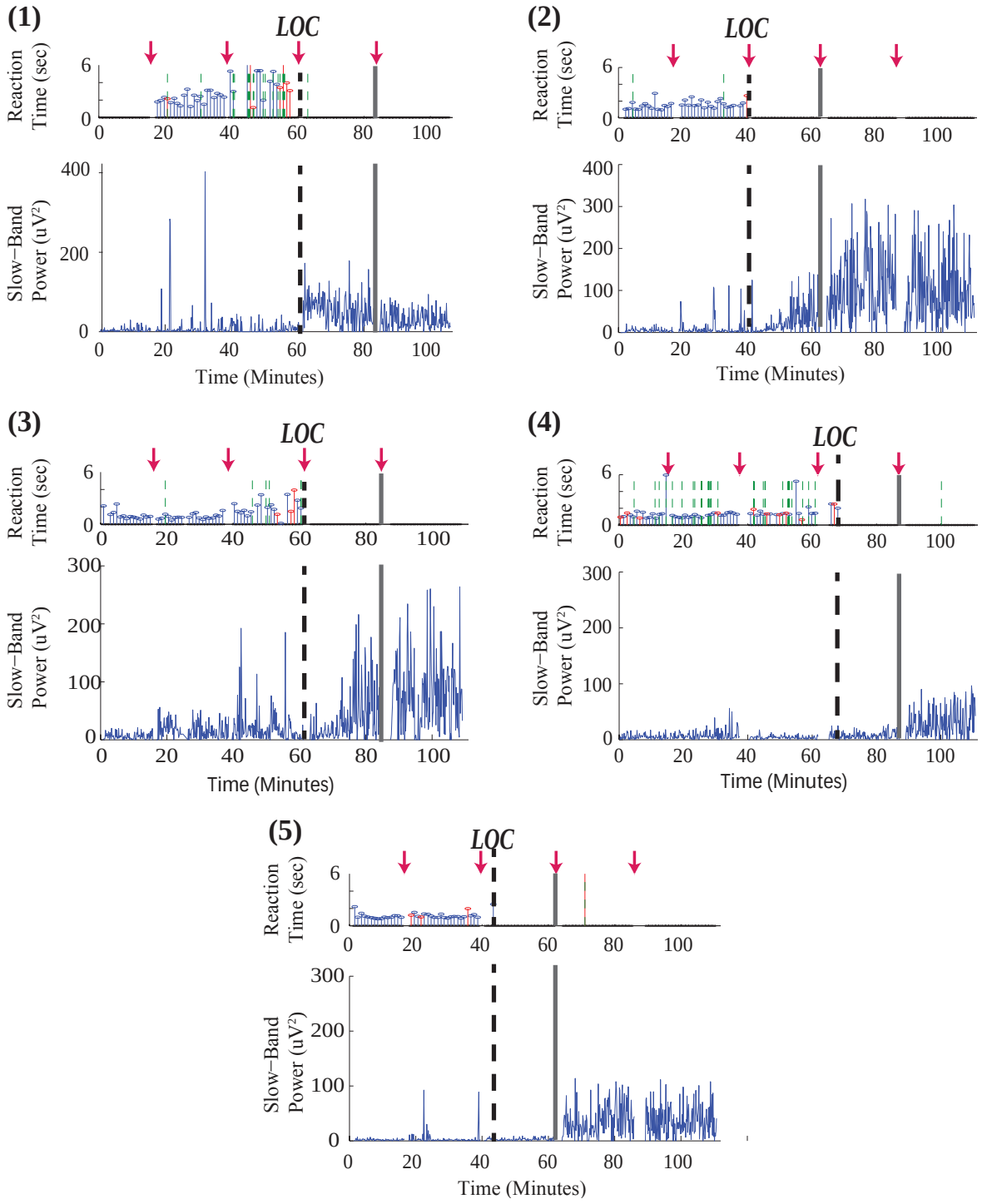

Figure 4-6: Slow-Band Power Across Behavioral States for 5 Subjects. Each panel corresponds to one subject. Panel numbering corresponds to subject numbers used in text. In a given panel, the red arrows denote time points corresponding to drug dose increments and top subplots show behavioral responses. The dashed black lines mark LOC, defining transition into unconsciousness (LOC+1). The solid gray lines mark transition into profound unconsciousness (LOC+2 and beyond). The bottom subplots show evolution of the band powers ($0.1 - 1$ Hz) through behavioral states. We note that Fig. 4-3-Fig. 4-4 show spectrograms from Subject 2.

4.3.3 EEG Phase-Amplitude Coupling Marks Transition in fMRI Responses During Unconsciousness

EEG modulograms were computed to quantify coupling between the phase of the slow oscillation and the amplitude of the alpha oscillation, and plotted alongside the fMRI results. Fig. 4-7 and Fig. 4-8 present the behavioral state (Panels *A*-*B*), EEG modulograms (Panel *C*), and fMRI-based regional dynamics (Panel *D*) for two subjects respectively. In each case, the modulogram plots are formatted consistently with those in Fig. 4-1. As the phase modulogram is best interpreted during periods when significant slow and frontal alpha range oscillations are present in the EEG, it is most meaningful to focus on the modulogram during the unconscious states.

In these 2 subjects, during profound levels of unconsciousness (LOC+2), we see a significant peak-max modulation pattern, where the peak of the slow-band oscillation modulates the alpha-band power. In this profoundly unconscious state, as previously noted, the fMRI-based auditory responses evidence a transition in corticothalamic dynamics – from suppressed sensory transmission into a state of intense cortical suppression and increased subcortical activation. Thus, the EEG-fMRI association suggests that the peak-max pattern may be a marker for a state of profound cortical inhibition and subcortical reactivation. As peak-max modulation marks a behaviorally non-apparent profoundly unconscious state, these preliminary associations with specific brain network dynamics offer a means to better interpret this EEG signature.

These results remain to be generalized across subjects and channels, however, we note that the peak-max modulation pattern is not limited by channel location. Fig. 4-7 shows the modulogram for an occipital channel on subject 2 while Fig. 4-8 shows the modulogram for a frontal channel on subject 4, indicating that the peak-max pattern may correspond to a widespread brain state. These modulation trends are similar to those seen outside the scanner [91, 150, 214] - with the exception that these data do not reveal a significant trough-max signature precluding LOC. This may be due to differences in behavioral paradigm employed and/or the subject demographic studied (see § 4.4)

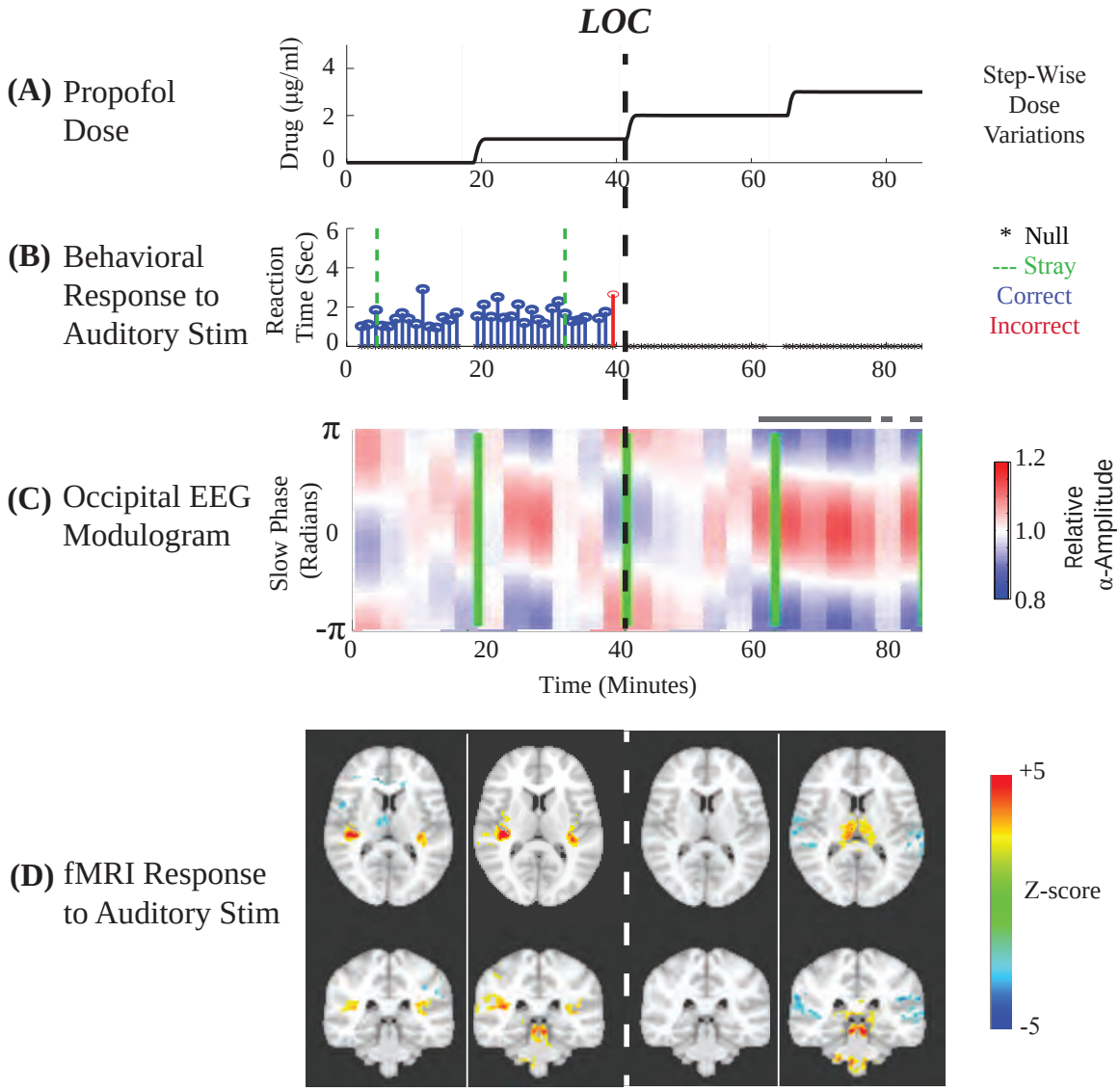

Figure 4-7: Occipital EEG Modulogram Alongside fMRI-based Regional Dynamics During Induction of Propofol Anesthesia. (*A-B*) Progressive propofol dosing and associated loss of behavioral response to auditory tone stimuli through the study. (*C*) Occipital EEG modulogram (O1 →P3) showing modulation of alpha power by slow wave peak (zero-phase). Gray bars represent significance ($p < 0.05$) by permutation test. *(A-C)* Data are from same subject as in Fig. 4-4 (subject 2). (*D*) Group fMRI responses to the auditory stimuli. (*C-D*) Peak-max EEG pattern during profound unconsciousness (LOC+2) is associated with state of intensified cortical inhibition and increased activation in subcortical regions on fMRI.

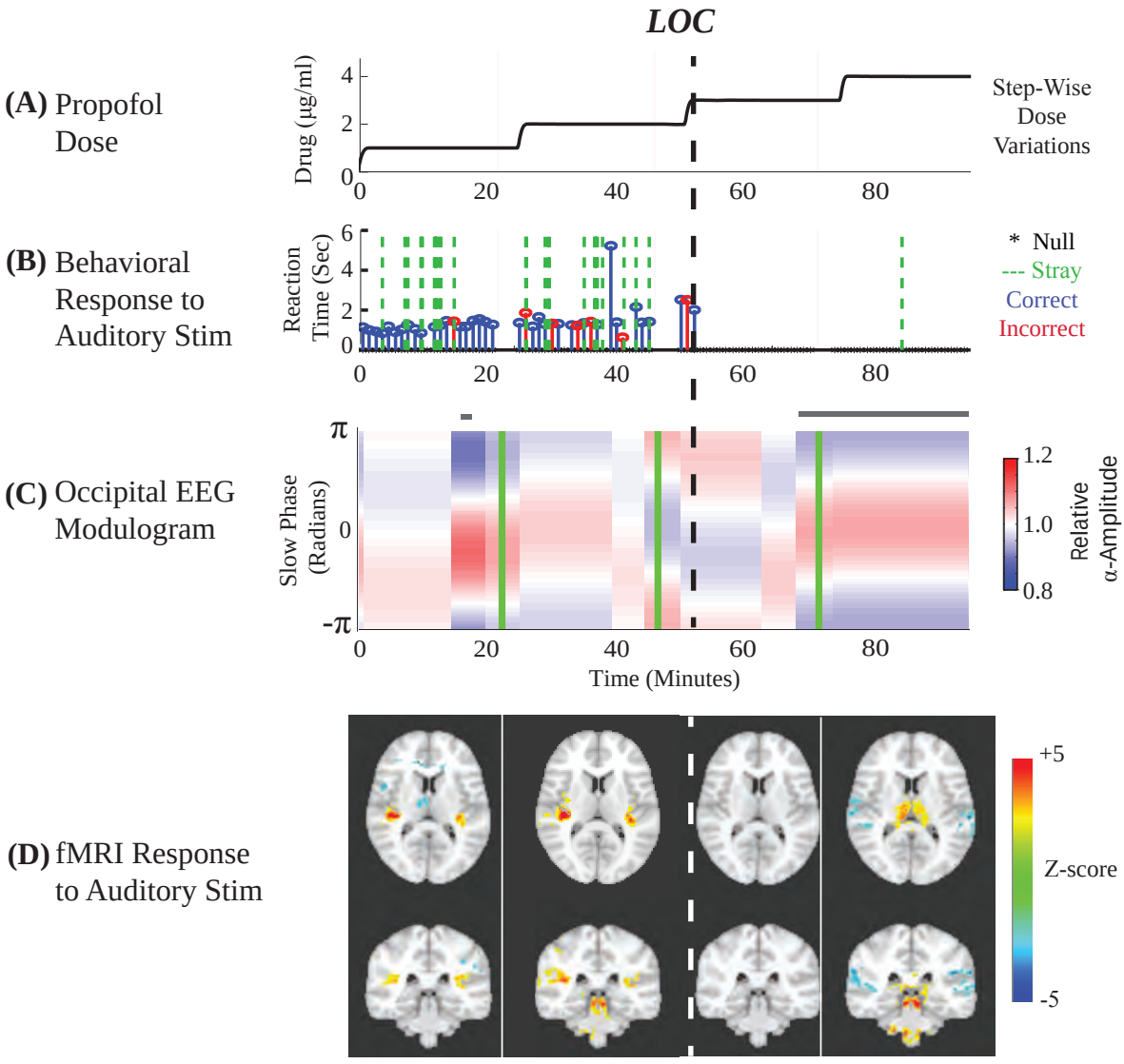

Figure 4-8: Frontal EEG Modulogram Alongside fMRI-based Regional Dynamics During Induction of Propofol Anesthesia. (*A-B*) Progressive propofol dosing and associated loss of behavioral response to auditory tone stimuli through the study. (*C*) Frontal EEG modulogram (F4 →Fp2) showing modulation of alpha power by slow wave peak (zero-phase). Gray bars represent significance ($p < 0.05$) by permutation test. *(A-C)* Data are from subject 4 of Fig. 4-5-Fig. 4-6. (*D*) Group fMRI responses to the auditory stimuli. (*C-D*) Peak-max EEG pattern during profound unconsciousness (LOC+2) is associated with state of intensified cortical inhibition and increased activation in subcortical regions on fMRI.

4.4 Discussions

In summary, we have related time-frequency EEG signatures marking changes in consciousness under propofol general anesthesia with simultaneous auditory fMRI responses across the brain. We have presented three main findings. First, the occipital alpha oscillations are associated with intact baseline processing of non-visual stimuli. Second, the onset of frontal alpha oscillations (alpha anteriorization) and widespread slow oscillations corresponds to suppression in primary sensory transmission at clinical LOC. Third, the onset of the peak-max phase-amplitude modulation pattern marks a state of intense cortical inhibition and subcortical reactivation that develops during profound unconsciousness well after clinical LOC. These findings associate ongoing EEG oscillatory dynamics with specific fMRI-based changes in sensory processing pathways and networks under propofol anesthesia. As such, they enhance understanding of functional and cognitive states underlying anesthesia EEG oscillations, and enable better interpretation of clinical EEG signatures demarcating states of consciousness under anesthesia.

4.4.1 Alpha Oscillations and Sensory Processing Under General Anesthesia

Our results elucidate roles of alpha-range EEG dynamics in modulating primary sensory responses. While occipital alpha oscillations are known to correspond to reduced visual processing, enhancement of processes within attentional focus, and filtering out of irrelevant information [25, 28, 239], their influence on activity in non-visual sensory pathways is sparsely studied. Our results provide evidence of intact responsiveness in non-visual sensory pathways during the baseline occipital alpha state, and thus are consistent with the belief that alpha oscillations have topographic specificity associated with sensory modality. Further, our results enhance understanding of how anesthesia-induced frontal alpha oscillations influence routing of sensory stimuli. A recent study used EEG source modeling to link propofol-induced frontal alpha activity with maintenance of early cortical sensory response and suppression of late cortical sensory response [221]. Our analysis extends these observations by associating frontal EEG alpha activity with suppression of fMRI-based sensory responses – to suggest suppressed routing of stimuli for both early and late stages of sensory processing across subcortical and cortical structures. This evidence for suppression through the auditory processing hierarchy suggests a possible effect on the ascending sensory pathway. These findings are consistent with previous work showing that the related state of spindling in sleep disrupts primary responses to auditory stimuli [77].

4.4.2 Slow Oscillations and Sensory Processing Under General Anesthesia

Furthermore, our results provide evidence linking EEG slow waves with altered stimulus response in networks across the brain. Anesthesia-induced slow waves have been shown to reflect fragmentation of cortico-cortical networks [214], but their association with network dynamics involving subcortical structures has not been resolved. We find that (a) the onset of EEG slow oscillations corresponds to loss of primary thalamocortical sensory response, and (b) the coupling of slow wave peaks

and alpha-band power in deeply anesthetized states is associated with intense cortical inhibition and subcortical reactivation. The intense cortical inhibition might be associated with a loss of descending inhibition in pathways projecting down from cortex to subcortical structures causing a reactivation in thalamus and brainstem [240]. Thus, the peak-max modulation pattern marking profound unconsciousness may correspond to a loss of descending corticothalamic inhibition. A recent propofol EEG-fMRI study associated EEG slow wave saturation in deep unconsciousness with loss of primary thalamocortical sensory response on fMRI [232]. By contrast, our data suggest (a) that thalamocortical isolation from sensory stimuli may be associated with onset of slow waves, thus precluding slow wave saturation, and (b) that profound unconsciousness is associated with a peak-max pattern corresponding to reactivation of the thalamus instead of reduction in thalamic activity. Whether these distinctions are due to differences in stimuli analyzed (tones in our case versus verbal stimuli in theirs) remains to be determined. Further, our EEG analyses in the scanner provide a richer set of EEG dynamics to associate with fMRI responses, allowing more nuanced categorization of brain states during profound unconsciousness.

Taken together, our results suggest that the onset of propofol alpha and slow oscillations at the transition into LOC is associated with suppression of ascending sensory processing pathways. On the other hand, the peak-max modulation pattern at more profoundly unconscious states is likely to be associated with stimulus-induced changes in descending corticothalamic pathways, wherein subcortical reactivation may be a readout of the degree of cortical inhibition.

4.4.3 Relevance for Brain State Monitoring Under General Anesthesia

As our results associate clinical EEG signatures with specific changes in pathways and networks involved in sensory processing, they provide a means to better interpret these clinical signatures. This has implications for brain state monitoring during general anesthesia. Common EEG monitors for depth of anesthesia provide indices that are not clearly connected to behavioral states or brain dynamics [150, 241]. As such, neurophysiologically based standards for brain state monitoring during anesthesia are desired to prevent complications of intraoperative awareness or recall [242]. Our results add to accumulating evidence on neurophysiologically based EEG signatures for monitoring states of profound unconsciousness [91, 150, 214, 232]. Specifically, our findings allow interpretation of the onset of slow and frontal alpha EEG oscillations as markers of decreased primary sensory response. Further, our results suggest that onset of the peak-max modulation pattern in profound unconsciousness marks the development of a distinct functional brain state - one that is characterized by intense cortical inhibition in response to sensory stimuli and coincidental subcortical reactivation. Recent reports have suggested that the EEG peak-max modulation pattern may correspond to a state where perceptual awareness is unlikely [91, 150]. Our results add dimension to these observations, by associating the peak-max modulation pattern with distinct fMRI-based brain dynamics. Further, we anticipate that the association of specific brain state changes with clinical EEG signatures may one day help anesthesiologists administer precise doses to target desired brain states. This could help prevent postoperative complications due to excessive dosing [243].

4.4.4 Future Directions

There remain limitations in the current study that can be addressed in the future. First, while EEG signatures recovered in this study qualitatively match trends seen in studies outside the scanner [91, 150, 213, 221], quantitative comparisons must be tempered by some nuances. The subjects in this study were $50-60\,\mathrm{y/o}$, an age range associated with lower alpha-band power than that seen in younger subjects [244, 245], the latter comprising the demographic typically studied outside the scanner. This also suggests a possible explanation for the lack of a significant trough-max state in our phase amplitude coupling results (Fig. 4-7-Fig. 4-8). Future studies may seek to recruit younger demographics with the required screening criteria for EEG-fMRI studies of general anesthesia. Second, the fMRI analysis was segmented by behavioral state and performed independently of the EEG data. Thus EEG-fMRI associations were purely based on alignment in time. This may not significantly affect results as the EEG changes and behavioral states are typically tightly linked. However, a future revision should reproduce these analyses by using EEG signatures in lieu of behavioral levels to segment the fMRI analyses.

Although our analysis of time-frequency EEG signatures obtained in the MRI scanner (characterizations spanning frequency bands and channels, and analyses of complex coupled oscillations) has not been previously matched in detail and scope, difficulties remain to be overcome. Specifically, computing phase-amplitude modulograms remains challenging given the short discontiguous format of EEG data acquired within the interleaved EEG-fMRI paradigm. Accurately computing slow phase for $3-5\,\mathrm{second}$ segments of data using frequency domain methods like filtering or Hilbert transforms is difficult, as these methods typically require data windows spanning a few minutes. Thus, in many cases, although time series EEG estimates visually appear to have phase-amplitude coupling, the modulograms and in turn statistical analyses of the modulograms do not adequately represent strength of coupling. Future efforts should seek to record longer EEG windows between fMRI volumes and ensure removal of gradient artifacts during fMRI scanning for long continuous EEG data segments. Further, as the autoregressive model used in the ballistocardiogram artifact removal does not encode any coupling between frequencies, the EEG estimates after removal may not exhibit appropriate coupling. That we see significant phase-amplitude coupling regardless attests to the strength of our analysis; however, future efforts could use a bispectral model in lieu of the autoregressive model to explicitly test cross-frequency coupling. Finally, bootstrap methods for analyzing significance of phase amplitude coupling should be evolved to improve statistical power over the current permutation testing approach. Despite these shortcomings, we note that our results, constituting the first report of EEG cross-frequency oscillatory coupling in the MRI scanner, follow general trends seen outside the scanner [91, 150]. Improvements with evolution of cross-frequency analysis methods suited to EEG-fMRI studies can be anticipated.

Finally, while our results suggest interplay between ongoing EEG oscillations and auditory fMRI responses (evoked by the tone stimuli), the extent to which they explain or relate to proposed mechanisms underlying genesis of the EEG oscillations [36, 199, 246, 247] is unclear. Future analyses

may distinguish fMRI responses to the EEG oscillatory powers from fMRI responses to the auditory stimulus. One way to do this is to pose separate fMRI analyses wherein the oscillatory power series forms the primary regressor and the stimulus sequence constitutes regressors of no interest, or vice versa. These approaches can inform future empirical studies of brain dynamics associated with the genesis of individual and coupled propofol EEG oscillations.

Chapter 5

Conclusions and Outlook

5.1 Thesis Summary

In this thesis, we have developed approaches and algorithms to enhance spatiotemporal imaging of human brain function. We have focused on challenges limiting two multimodal approaches for spatiotemporal imaging: (a) simultaneous EEG-fMRI, and (b) source imaging using M/EEG and anatomic MRI.

In Chapter 2, we developed an approach to remove EEG-fMRI ballistocardiogram artifacts. First, we defined a mathematical harmonic basis for these artifacts. This allowed us to frame the artifact removal problem as one of model-based parameter estimation without requiring templates based on cardiac or motion reference signals. Next, we developed a real-time local likelihood algorithm to estimate model parameters and separate MR-related artifact from brain-generated EEG signals. Then, we demonstrated effective recovery of neurophysiologic oscillatory and evoked EEG signatures in challenging SNR conditions within the MRI scanner. We further showed that the model and algorithm afford significant improvements over existing reference-based techniques, thus enabling improved sensitivity and specificity for resolving EEG dynamics in the MRI scanner.

In Chapter 3, we developed an approach to extend electromagnetic source imaging with M/EEG and anatomic MRI to subcortical structures. To overcome challenges in resolving deep subcortical sources that contribute low amplitude signals to surface M/EEG sensors, we focused on amplitude-insensitive information in M/EEG field patterns. First, we identified robust distinctions in M/EEG field patterns generated by sparse sets of subcortical and cortical sources. Next, we translated these insights into a practical and efficient solution for subcortical source imaging. Specifically, we developed an efficient hierarchical subspace pursuit algorithm to estimate neural currents in subcortical structures. Then, we demonstrated efficacy for localizing thalamic and brainstem contributions in challenging evoked response test cases, thus extending the spatial span of source imaging studies to include millisecond-scale dynamics involving subcortical structures.

In Chapter 4, we illustrated the potential of these advances in spatiotemporal imaging for enhanced studies of brain dynamics. Specifically, we focused on the study of systems-level brain dynamics underlying loss of consciousness under general anesthesia. We performed time-frequency analyses on EEG obtained in the MRI scanner during propofol anesthesia, and associated the resulting oscillatory EEG dynamics with simultaneous auditory fMRI responses across brain regions. Our findings provide novel insights into the roles of propofol-induced alpha and slow oscillations in modulating sensory response under progressively deep stages of anesthesia. These results provide neurophysiologic evidence to help better interpret functional brain states corresponding to clinical EEG signatures observed during propofol anesthesia, even when no behavioral changes are apparent.

Overall, this work enables enhanced resolution of spatiotemporal brain dynamics with human multimodal imaging, and opens opportunities for new lines of investigation in cognitive and clinical neuroscience.

5.2 Future Directions

5.2.1 Approaches for Biomedical Signal Processing

We discuss perspectives uniting approaches used in this thesis to overcome fundamental biophysical and signal processing challenges. For each of the core problems, we started with a systematic consideration of the physiology, anatomy and physics underlying the data. This helped identify natural structure and uncover non-apparent information present in the data. We then utilized this prior information to develop statistically principled estimation algorithms and overcome limitations in the field.

First, for the EEG-fMRI artifact removal problem (Chapter 2), much of the literature uses reference signals within black-box approaches. By contrast, we analyzed the EEG data in context of the physiology and physics underlying the artifacts, to identify a harmonic basis for the artifacts (Fig. 2-2). This then motivated the choice for a statistical model and harmonic regression algorithm, which, in turn, enables improved artifact removal and recovery of EEG dynamics in the MRI scanner. Second, for the source imaging problem (Chapter 3), much of the literature has neglected subcortical sources, as they generate low amplitude M/EEG signals. By contrast, we analyzed the M/EEG field patterns in context of the anatomy underlying the source space and the physics underlying the M/EEG recordings. Our analysis revealed robust distinctions between subcortical and cortical field patterns (Fig. 3-2-Fig. 3-3). This then motivated the choice of a hierarchical projection pursuit algorithm, which in turn, enables subcortical source estimation.

These case studies have implications for the analysis of biomedical data. Many biomedical signal processing problems typically have three sources of prior information. First, the physiology contains information regarding dynamics at play during the process being measured. Second, the anatomy

contains prior information. Third, measurement physics translates the anatomy and physiology underlying the process under study to the data being analyzed. While complete information about all priors is often lacking in practice, we find that systematic exploratory data analysis contextualized within this framework can often reveal surprisingly elegant information in the data. Further, starting with the structure in the problem rather than a technique motivates customized algorithmic choices, encourages explicit statement of models, assumptions or constraints, and is also amenable to goodness of fit analyses. For example, characterization of the sparsity constraints required to distinguish subcortical and cortical field maps motivated a hierarchical source space reduction procedure tailored to these constraints (§ 3.3). Finally, such approaches may offer exceptions to general theoretical limits that do not account for the specific prior information at hand. For example, it is known that short data segments cannot, in general, be used to obtain high spectral resolution. However, for the specific case where prior harmonic structure is known, the limits that apply are different from those for the general case; and a statistical procedure can be devised to obtain high resolution with short data segments (§ 2.5.2). While we have demonstrated specific applications for analysis of neural and imaging data, these general ideas are also relevant for analysis and interpretation of the increasingly large, complex and multimodal biomedical datasets emerging beyond neuroscience and imaging.

5.2.2 Techniques for Multimodal Neuroimaging

The algorithms developed in this thesis have implications for enhanced multimodal neuroimaging.

The framework in Chapter 2 has implications beyond ballistocardiogram artifact removal. First, by defining a mathematical basis for ballistocardiogram artifacts (Fig. 2-2), our work helps frame the analysis of EEG signals recorded in the MRI scanner within a model-based inference framework. This can help make quantitative statistically principled statements about a wide variety of EEG time-frequency features, beyond the oscillatory dynamics we have studied in this thesis. For example, other EEG time-frequency features can be modeled in lieu of the autoregressive forms, and estimated within a statistical framework similar to the one we have derived. Second, while common gradient artifact removal methods subtracting time-locked pulses are known to perform reliably, they sometimes leave residuals of the same order of magnitude as EEG features of interest [97, 102]. Because our algorithm can remove periodic artifacts in EEG, applying our algorithm during the fMRI acquisition periods, with a fundamental frequency corresponding to the gradient artifacts, can improve these residuals. Third, a quantifiable form for MR-related artifacts can be useful in the design of EEG electrodes or head restraint methods for reduced-artifact acquisition [248, 249]. As our algorithm quantifies the artifacts in parametric form across channel locations (§ 2.2.3-§ 2.3), it can enable better assessment and benchmarking of newer designs. Overall, these observations suggest broader implications of our approach and algorithm for enhanced EEG-fMRI acquisition and analysis.

The approach for subcortical source imaging (Chapter 3) has general implications for paradigms beyond the specific problem of resolving evoked subcortical activity. While we have demonstrated that a framework employing field pattern differences can estimate subcortical sources when spatial sparsity constraints are imposed, the general ideas apply to cases where activity may be spatially distributed but sparse in another basis. Sparsity in time, in frequency, or in combinations of time, frequency and spatial domains, are increasingly employed in source imaging approaches to explore the dynamical nature of cortical function [86, 190, 193–198], and thus have relevance for studies of subcortical function as well. We note that the class of projection pursuit algorithms and hierarchical reduction procedures proposed (§ 3.3) are well-suited to exploit sparsity in these other bases, thus suggesting feasibility of the approaches outlined for studies of widespread oscillatory and network dynamics across a range of paradigms. Beyond estimating subcortical sources, we have demonstrated that cortical solutions can be misled to explain subcortical components when the latter are not appropriately accounted for (Fig. 3-10). Thus, interpretation of typical source estimation results that use only cortical source spaces must account for this caveat. Overall, the ideas we have put forth have implications for source imaging across experimental paradigms, and for estimation of both subcortical and cortical contributions to M/EEG recordings. Future work should validate localization accuracy with direct intracranial recordings in subcortical structures like hippocampus or basal ganglia alongside surface M/EEG recordings. A behavioral paradigm that elicits predictable changes in subcortical activity would be fruitful for such an endeavor.

Together, Chapter 2-Chapter 3 have implications for joint electromagnetic source imaging of brain dynamics with combined electromagnetic and MRI information. Specifically, improved signal-to-noise ratio for EEG in the MRI scanner can enhance fMRI-informed EEG source imaging. fMRI-informed electromagnetic source imaging is commonly performed with separately acquired EEG and fMRI owing to poor EEG signal quality [104, 165, 166]. Reduced EEG artifacts across channels with varied SNRs can allow electrical source imaging to be performed in an integrated framework with simultaneously acquired EEG-fMRI data. Further, our hierarchical source space reduction procedure suggests a framework to enhance source imaging by incorporating two types of MRI-based prior information. First, having concurrently obtained fMRI with clean EEG offers spatial priors to better inform the hierarchical reduction of source spaces for our subcortical source estimation approach. As we have shown that EEG field patterns from subcortical and cortical sources exhibit relatively weak distinctions (Fig. 3-13), simultaneously acquired fMRI-priors can inform and improve subcortical source estimation based on EEG data. Second, anatomic information from diffusion tensor imaging (DTI) [250] can be incorporated to inform the subcortical source estimation. Specifically, it may be possible to locate active cortical regions relevant to a given M/EEG measurement, identify a set of structurally connected subcortical regions, form a reduced source space of relevant and structurally connected cortical and subcortical regions, and then localize sources within this reduced space to explain the data. Overall, our work enables improved EEG SNR in the MRI scanner and develops a source space reduction procedure where MRI priors can be integrated, and thus has implications for enhanced joint electromagnetic source imaging.

5.2.3 Studies of Brain Dynamics

The algorithms developed in this thesis enhance the ability to relate fast timescale electrophysiologic measures of neuronal activity on M/EEG with their underlying regional dynamics.

Chapter 2 developed techniques to improve analysis of EEG recorded in the scanner, thus allowing better association between EEG time-frequency signatures and spatially resolved, albeit slower, measures of regional dynamics from fMRI. The analysis of EEG time-frequency signatures in the scanner has been largely focused on evoked potentials and epileptic spikes (§ 2.7). We extend the scope of these analyses to detailed time-frequency features relevant for the study of oscillatory rhythms. Specifically, we showed the ability to recover oscillatory dynamics on EEG across frequencies and channel locations that are typically associated with poor SNR (Fig. 2-6). Further, in Chapter 4, we provided the first report of coupling between EEG oscillations within the MRI scanner (Fig. 4-7-Fig. 4-8), and demonstrated applications for enhanced studies of brain dynamics during anesthesia. Thus, this work opens opportunities for improved studies of spatiotemporal dynamics underlying oscillatory rhythms that are relevant in a large number of states including sleep, attention, memory and perception (§ 1.2).

Chapter 3 demonstrated feasibility to use M/EEG and anatomical MRI for estimation of fast timescale regional dynamics involving subcortical structures. Existing source imaging techniques have largely focused on cortical sources (§ 3.1), and neglected subcortical dynamics that may be manifest in the M/EEG data. We include subcortical sources within the purview of M/EEG source imaging, and thus expand the scope of brain dynamics that can be studied using these techniques. Further, our field pattern analyses suggest that multimodal M/EEG recordings may be more informative for subcortical source estimation than unimodal recordings (§ 3.6, Fig. 3-13), and thus inform design of experimental paradigms to probe subcortical dynamics. Thus, this work has the potential to impact studies of dynamics involving subcortical regions such as thalamus, hippocampus, amygdala and basal ganglia, which are thought to have important roles in arousal, memory, emotion and motor control respectively [107, 110, 111].

Taken together, the two techniques developed in this thesis have implications for studies of regional brain dynamics spanning slow and fast timescales. One application area is in studies of the neurophysiology associated with the BOLD-fMRI response [163, 164]. As our BCG removal algorithm enables enhanced EEG time-frequency analyses across channels, it expands opportunities to link fMRI-based functional connectivity (slow timescales) with neurophysiologic signatures like EEG-based coherence (fast timescales) across brain networks. Further, the source estimation algorithm can be used to better understand similarities and differences between BOLD-fMRI responses (slow timescales) and M/EEG source current estimates (fast timescales) obtained from recordings under the same conditions.

Another application area is in studies of oscillatory brain dynamics across spatial scales and expanded brain networks. The study of brain dynamics during anesthesia (Chapter 4) provides a

good example to illustrate this opportunity. First, there is growing interest in how propofol alters functional connectivity across brain regions assessed by fMRI [53, 228–231]. With the technique developed in Chapter 2, these analyses could be expanded to relate EEG coherence (fast timescales) with fMRI-based connectivity (slow timescales). Second, while the simultaneous EEG-fMRI results associate coupling of propofol EEG oscillations with possible dynamics in a descending corticothalamic network (§ 4.4), they do not explain how dynamical interactions between cortical and thalamic regions may generate the coupling between multiple EEG oscillations. Such investigations need faster timescales for regional dynamics than fMRI can provide, and thus can be performed with the expanded subcortical source estimation framework. Thus, future work applying source estimation on high-density EEG recordings obtained during propofol anesthesia could help uniquely address these emergent thalamocortical oscillatory phenomena. One starting point may be to focus on thalamocortical dynamics underlying the widely studied alpha oscillations, and assess how these dynamics are modulated by propofol anesthesia. Such efforts would (a) add empirical confirmation and anatomic correlates to predictive cellular-level biophysical models [36, 199], and (b) relate results from electrophysiologic recordings in animals [240, 246, 247, 251–254] with changes in human behavioral states.

5.2.4 Clinical Neuroscience Studies

Beyond high spatiotemporal resolution for brain dynamics, our work has specific implications for clinical neuroscience.

In Chapter 4, we demonstrated associations between time-frequency EEG signatures and the underlying fMRI-based changes in sensory processing during general anesthesia. These EEG-fMRI associations allow improved interpretation of functional brain states underlying EEG dynamics during general anesthesia, even when no behavioral changes are apparent (§ 4.3), and thus illustrate two specific opportunities for clinical neuroscience studies.

First, EEG signatures can be assessed in individual subjects, while fMRI maps are typically presented as group averages (owing to low SNR and challenges in quantitative interpretation of BOLD signals, § 1.1). Thus, enhanced correlations between EEG and fMRI offer a means to obtain individually measurable markers of regional brain dynamics. This has implications for monitoring brain state and function in a patient and condition specific manner. Second, the cognitive and functional brain states associated with many M/EEG signatures cannot be completely categorized using behavioral measures alone. Thus, there is a need to associate M/EEG signatures with regional brain dynamics to better interpret the underlying functional brain state. Thus, spatiotemporal brain imaging has the potential to advance objective measures with predictive or diagnostic utility for behaviorally non-apparent changes in brain function – for e.g., changes in arousal, awareness, perception, emotion and memory in states of coma, hallucination, pain, depression and dementia. With improved rigor and robustness of EEG-fMRI and electromagnetic source imaging, we envision enhanced opportunities for these goals in clinical neuroscience.

Bibliography

[1] Matti S. Hamalainen. *Lecture notes on basics of MEG/EEG source modeling, Martinos Center Multimodal Imaging Course, Massachusetts General Hospital.* 2013.

[2] *Neuronal synapse: http://commons.wikimedia.org/wiki/File:Neuronal_Synapse.jpg.* 2011.

[3] Dale Purves, George J Augustine, David Fitzpatrick, William C Hall, Anthony-Samuel LaMantia, James O McNamara, and Leonard E White. *Neuroscience.* Sinauer Associates, fourth edition, 2008.

[4] Ernst Neidermeyer and Fernando Lopes da Silva, editors. *Electroencephalography: basic principles, clinical applications and related fields.* Lippincott Williams & Wilkins, sixth edition, 2005.

[5] Paul L Nunez and Ramesh Srinivasan. *Electric fields of the brain: the neurophysics of EEG.* Oxford University Press, USA, second edition, 2006.

[6] Matti S. Hamalainen. Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–505, April 1993.

[7] David Cohen. Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents. *Science*, 161:784–786, August 1968.

[8] David Cohen. Magnetoencephelography: detection of the brain's electrical activity with a superconducting magnetometer. *Science*, 175:664–666, February 1972.

[9] David Cohen. The MEG made ridiculously simple. *Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA*, 2009.

[10] Miles Wernick and John Aarsvold, editors. *Emission Tomography: the fundamentals of PET and SPECT.* Elsevier Academic Press, first edition, 2004.

[11] David Townsend, Peter Valk, Michael Maisey, and Dale Bailey. *Positron Emission Tomography.* Springer-Verlag London Limited, first edition, 2005.

[12] Jacob Hooker. *Lecture notes on Positron Emission Tomography basics, Martinos Center Multimodal Imaging Course, Massachusetts General Hospital.* 2013.

[13] Paul L. Nunez. *Neocortical dynamics and human EEG rhythms.* Oxford University Press, USA, first edition, 1995.

[14] Seiji Ogawa, Tso-Ming Lee, A R Kay, and D W Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–72, December 1990.

[15] Jack W Belliveau, David N Kennedy, Robert C McKinstry, Bradley R Buchbinder, Robert M Weisskoff, Mark S Cohen, J M Vevea, Thomas J Brady, and Bruce R Rosen. Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, 254(5032):716–719, November 1991.

[16] Kenneth K Kwong, Jack W Belliveau, D A Chesler, I E Goldberg, Robert M Weisskoff, B P Poncelet, David N Kennedy, B E Hoppel, Mark S Cohen, and Robert Turner. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, 89(12):5675–9, June 1992.

[17] Nikos K Logothetis. What we can do and what we cannot do with fMRI. *Nature*, 453(7197):869–78, June 2008.

[18] Richard B Buxton. Interpreting oxygenation-based neuroimaging signals: the importance and the challenge of understanding brain oxygen metabolism. *Frontiers in Neuroenergetics*, 2(8):1–16, January 2010.

[19] Timothy L Davis, Kenneth K Kwong, Robert M Weisskoff, and Bruce R Rosen. Calibrated functional MRI: mapping the dynamics of oxidative metabolism. *Proceedings of the National Academy of Sciences*, 95(4):1834–9, February 1998.

[20] Theodore J Huppert, Monica S Allen, Solomon G Diamond, and David A Boas. Estimating cerebral oxygen metabolism from fMRI with a dynamic multicompartment Windkessel model. *Human Brain Mapping*, 30(5):1548–67, May 2009.

[21] Sylvain Baillet, John C Mosher, and Richard M Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, pages 14–30, November 2001.

[22] Hellmuth Obrig and Arno Villringer. Beyond the visible-imaging the human brain with light. *Journal of Cerebral Blood Flow and Metabolism*, 23(1):1–18, January 2003.

[23] Marco Ferrari and Valentina Quaresima. A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage*, 63(2):921–35, November 2012.

[24] Emery N Brown, Ralph Lydic, and Nicholas D Schiff. General anesthesia, sleep and coma. *New England Journal of Medicine*, 363(27):2638–2650, December 2010.

[25] Wolfgang Klimesch. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29(2-3):169–95, April 1999.

[26] Laura Lee Colgin, Tobias Denninger, Marianne Fyhn, Torkel Hafting, Tora Bonnevie, Ole Jensen, May-Britt Moser, and Edvard I Moser. Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature*, 462(7271):353–7, November 2009.

[27] Andreas K Engel and Pascal Fries. Beta-band oscillations-signalling the status quo? *Current Opinion in Neurobiology*, 20(2):156–65, April 2010.

[28] Wolfgang Klimesch. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, 16(12):606–17, December 2012.

[29] Laura Lee Colgin. Mechanisms and functions of theta rhythms. *Annual Review of Neuroscience*, 36:295–312, July 2013.

[30] Catherine E Kerr, Matthew D Sacchet, Sara W Lazar, Christopher I Moore, and Stephanie R Jones. Mindfulness starts with the body: somatosensory attention and top-down modulation of cortical alpha rhythms in mindfulness meditation. *Frontiers in Human Neuroscience*, 7(12):1–15, February 2013.

[31] Marta Kutas and Kara D Federmeier. Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12):463–470, December 2000.

[32] Nathan E Crone, Dana Boatman, Barry Gordon, and Lei Hao. Induced electrocorticographic gamma activity during auditory perception. *Clinical Neurophysiology*, 112(4):565–582, April 2001.

[33] Thomas Andrillon, Yuval Nir, Richard J Staba, Fabio Ferrarelli, Chiara Cirelli, Giulio Tononi, and Itzhak Fried. Sleep spindles in humans: insights from intracranial EEG and unit recordings. *The Journal of Neuroscience*, 31(49):17821–17834, December 2011.

[34] Marta I Garrido, James M Kilner, Klaas E Stephan, and Karl J Friston. The mismatch negativity: a review of underlying mechanisms. *Clinical Neurophysiology*, 120(3):453–63, March 2009.

[35] Risto Näätänen, Teija Kujala, Kairi Kreegipuu, Synnöve Carlson, Carles Escera, Torsten Baldeweg, and Curtis Ponton. The mismatch negativity: an index of cognitive decline in neuropsychiatric and neurological diseases and in ageing. *Brain*, 134(Pt 12):3435–53, December 2011.

[36] Sujith Vijayan and Nancy J Kopell. Thalamic model of awake alpha oscillations and implications for stimulus processing. *Proceedings of the National Academy of Sciences*, 109(45):18553–8, November 2012.

[37] Michael Murphy, Brady A Riedner, Reto Huber, Marcello Massimini, Fabio Ferrarelli, and Giulio Tononi. Source modeling sleep slow waves. *Proceedings of the National Academy of Sciences*, 106(5):1608–13, February 2009.

[38] Timothy J Buschman, Eric L Denovellis, Cinira Diogo, Daniel Bullock, and Earl K Miller. Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron*, 76(4):838–46, November 2012.

[39] Markus Siegel, Tobias H Donner, and Andreas K Engel. Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews Neuroscience*, 13(2):121–34, February 2012.

[40] Joerg F Hipp, David J Hawellek, Maurizio Corbetta, Markus Siegel, and Andreas K Engel. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature Neuroscience*, 15:884–890, May 2012.

[41] Mircea Steriade, David A McCormick, and Terrence J Sejnowski. Thalamocortical oscillations in the sleeping and aroused brain. *Science*, 262(5134):679–85, October 1993.

[42] Vladyslav V. Vyazovskiy, Umberto Olcese, Erin C. Hanlon, Yuval Nir, Chiara Cirelli, and Giulio Tononi. Local sleep in awake rats. *Nature*, 472(7344):443–447, April 2011.

[43] Laure Buhry, Amir H Azizi, and Sen Cheng. Reactivation, replay, and preplay: how it might all fit together. *Neural Plasticity*, 2011:203462, January 2011.

[44] Stephen A Engel, Gary H Glover, and Brian A Wandell. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, 7(2):181–92, March 1997.

[45] Adrian M Owen, Martin R Coleman, Melanie Boly, Matthew H Davis, Steven Laureys, and John D Pickard. Detecting awareness in the vegetative state. *Science*, 313(5792):1402, October 2006.

[46] Mark D' Esposito, Brad R Postle, D Ballard, and J Lease. Maintenance versus manipulation of information held in working memory: an event-related fMRI study. *Brain and Cognition*, 41(1):66–86, October 1999.

[47] Anthony D Wagner, Daniel L Scacter, Michael Rotte, Wilma Koutstaal, Anat Maril, Anders M Dale, Bruce R Rosen, and Randy L Buckner. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, 281(5380):1188–1191, August 1998.

[48] Hans C Breiter, Nancy L Etcoff, Paul J Whalen, William A Kennedy, Scott L Rauch, Randy L Buckner, Monica M Strauss, Steven E Hyman, and Bruce R Rosen. Response and habituation of the human amygdala during visual processing of facial expression. *Neuron*, 17(5):875–87, December 1996.

[49] Richard J Davidson. Anxiety and affective style: role of prefrontal cortex and amygdala. *Biological Psychiatry*, 51(1):68–80, January 2002.

[50] Hans C Breiter, Randy L Gollub, Robert M Weisskoff, David N Kennedy, Nikos Makris, Joshua D Berke, Julie M Goodman, Howard L Kantor, David R Gastfriend, Jonn P Riorden, R Thomas Mathew, Bruce R Rosen, and Steven E Hyman. Acute effects of cocaine on human brain activity and emotion. *Neuron*, 19(3):591–611, October 1997.

[51] Dominic H Ffytche, Robert J Howard, Michael J Brammer, Anthony David, Peter Woodruff, and Steven Williams. The anatomy of conscious vision: an fMRI study of visual hallucinations. *Nature Neuroscience*, 1(8):738–42, December 1998.

[52] Tor D Wager, Lauren Y Atlas, Martin A Lindquist, Mathieu Roy, Choong-Wan Woo, and Ethan Kross. An fMRI-based neurologic signature of physical pain. *The New England Journal of Medicine*, 368(15):1388–97, April 2013.

[53] Xiaolin Liu, Kathryn Lauer, Douglas Ward, Shi-Jiang Li, and Anthony G Hudetz. Differential effects of deep sedation with propofol on the specific and nonspecific thalamocortical systems. *Anesthesiology*, 118(1):59–69, January 2013.

[54] Marcus E Raichle, Ann M MacLeod, Abraham Z Snyder, William J Powers, Debra A Gusnard, and Gordon L Shulman. A default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(2):676–82, January 2001.

[55] Randy L Buckner, Jessica R Andrews-Hanna, and Daniel L Schacter. The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124:1–38, March 2008.

[56] Bharat Biswal, F Zerrin Yetkin, Victor M Haughton, and James S Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34(9):537–541, June 1995.

[57] Michael D Greicius, Gaurav Srivastava, Allan L Reiss, and Vinod Menon. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proceedings of the National Academy of Sciences*, 101(13):4637–42, March 2004.

[58] Pieter Guldenmund, Athena Demertzi, Pierre Boveroux, Mélanie Boly, Audrey Vanhaudenhuyse, Marie-Aurélie Bruno, Olivia Gosseries, Quentin Noirhomme, Jean-François Brichant, Vincent Bonhomme, Steven Laureys, and Andrea Soddu. Thalamus, brainstem and salience network connectivity changes during propofol-induced sedation and unconsciousness. *Brain connectivity*, 3(3):273–85, January 2013.

[59] Bruce R Rosen and Robert L Savoy. fMRI at 20: has it changed the world? *NeuroImage*, 62(2):1316–24, August 2012.

[60] John R Ives, Steven Warach, Franz Schmitt, RR Edelman, and Donald L Schomer. Monitoring the patient's EEG during echo planar MRI. *Electroencephalography and Clinical Neurophysiology*, 87:417–420, August 1993.

[61] Robin I Goldman, John M Stern, Jerome Engel, and Mark S Cohen. Acquiring simultaneous EEG and functional MRI. *Clinical Neurophysiology*, 111(11):1974–80, November 2000.

[62] Louis Lemieux, Afraim Salek-Haddadi, Oliver Josephs, Philip J Allen, Nathan Toms, Catherine Scott, Karsten Krakow, Robert Turner, and David R Fish. Event-related fMRI with simultaneous and continuous EEG: description of the method and initial case report. *NeuroImage*, 14(3):780–7, September 2001.

[63] Patrick L Purdon, Eric T Pierce, Giorgio Bonmassar, John Walsh, P Grace Harrell, Jean Kwo, Daniel Deschler, Margaret Barlow, Rebecca C Merhar, Catherine M Mullaly, Mary Sullivan, Sharon Maginnis, Debra Skoniecki, Anne Higgins, and Emery N Brown. Simultaneous electroencephalography and functional magnetic resonance imaging of general anesthesia. *Annals of the New York Academy of Sciences*, 1157:61–70, March 2009.

[64] Robin I Goldman, John M Stern, Jerome Engel, and Mark S Cohen. Simultaneous EEG and fMRI of the alpha rhythm. *NeuroReport*, 13(18):2487–92, December 2002.

[65] Mathias Schreckenberger, Christian Lange-Asschenfeldt, Christian Lange-Asschenfeld, Matthias Lochmann, Klaus Mann, Thomas Siessmeier, Hans-Georg Buchholz, Peter Bartenstein, and Gerhard Gründer. The thalamus as the generator and modulator of EEG alpha rhythm: a combined PET/EEG study with lorazepam challenge in humans. *NeuroImage*, 22(2):637–44, June 2004.

[66] Kay Jann, T. Dierks, C. Boesch, M. Kottlow, W. Strik, and Thomas Koenig. BOLD correlates of EEG alpha phase-locking and the fMRI default mode network. *NeuroImage*, 45(3):903–916, April 2009.

[67] Zhongming Liu, Jacco A de Zwart, Bing Yao, Peter van Gelderen, Li-Wei Kuo, and Jeff H Duyn. Finding thalamic BOLD correlates to posterior alpha EEG. *NeuroImage*, 63(3):1060–9, November 2012.

[68] Stephen D Mayhew, Dirk Ostwald, Camillo Porcaro, and Andrew P Bagshaw. Spontaneous EEG alpha oscillation interacts with positive and negative BOLD responses in the visual-auditory cortices and default-mode network. *NeuroImage*, 76:362–372, March 2013.

[69] Jue Mo, Yuelu Liu, Haiqing Huang, and Mingzhou Ding. Coupling between visual alpha oscillations and default mode activity. *NeuroImage*, 68:112–8, March 2013.

[70] Rikkert Hindriks and Michel J A M van Putten. Thalamo-cortical mechanisms underlying changes in amplitude and frequency of human alpha oscillations. *NeuroImage*, 70:150–63, April 2013.

[71] Helmut Laufs, Karsten Krakow, P Sterzer, E Eger, A Beyerle, Afraim Salek-Haddadi, and Andreas Kleinschmidt. Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proceedings of the National Academy of Sciences*, 100(19):11053–8, September 2003.

[72] Helmut Laufs, John L Holt, Robert Elfont, Michael Krams, Joseph S Paul, Karsten Krakow, and Andreas Kleinschmidt. Where the BOLD signal goes when alpha EEG leaves. *NeuroImage*, 31(4):1408–18, July 2006.

[73] Donald L Schomer, Giorgio Bonmassar, F Lazeyras, M Seeck, A Blum, Kimitaka Anami, D Schwartz, Jack W Belliveau, and John R Ives. EEG-linked functional magnetic resonance imaging in epilepsy and cognitive neurophysiology. *Journal of Clinical Neurophysiology*, 17(1):43–58, January 2000.

[74] Giandomenico D Iannetti, C Di Bonaventura, P Pantano, A T Giallonardo, P L Romanelli, L Bozzao, Mario Manfredi, and G B Ricci. fMRI/EEG in paroxysmal activity elicited by elimination of central vision and fixation. *Neurology*, 58(6):976–9, March 2002.

[75] Manuel Schabus, Thien Thanh Dang-Vu, Genevieve Albouy, Evelyne Balteau, Melanie Boly, Julie Carrier, Annabelle Darsaud, Christian Degueldre, Martin Desseilles, Stephen Gais, Christophe Phillips, G Rauchs, C Schnakers, V Sterpenich, G Vandewalle, A Luxen, and Pierre Maquet. Hemodynamic cerebral correlates of sleep spindles during human non-rapid eye movement sleep. *Proceedings of the National Academy of Sciences*, 104(32):13164–9, August 2007.

[76] Thien Thanh Dang-vu, Manuel Schabus, Martin Desseilles, Genevieve Albouy, Melanie Boly, Annabelle Darsaud, Stephen Gais, Virginie Sterpenich, Gilles Vandewalle, Julie Carrier, Gustave Moonen, Evelyne Balteau, Christian Degueldre, Andre Luxen, Christophe Phillips, and Pierre Maquet. Spontaneous neural activity during human slow wave sleep. *Proceedings of the National Academy of Sciences*, 105(39):15160–15165, July 2008.

[77] Thien Thanh Dang-vu, Maxime Bonjean, Manuel Schabus, Mélanie Boly, and Annabelle Darsaud. Interplay between spontaneous and induced brain activity during human non-rapid eye movement sleep. *Proceedings of the National Academy of Sciences*, 108:37:15438–15443, 2011.

[78] Mélanie Boly, Vincent Perlbarg, Guillaume Marrelec, Manuel Schabus, Steven Laureys, Julien Doyon, Mélanie Pélégrini-Issac, Pierre Maquet, and Habib Benali. Hierarchical clustering of brain activity during human nonrapid eye movement sleep. *Proceedings of the National Academy of Sciences*, 109(15):5856–61, April 2012.

[79] Simultaneous EEG-fMRI: http://fmri.uib.no/.

[80] Philip J Allen, G Polizzi, Karsten Krakow, David R Fish, and Louis Lemieux. Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction. *NeuroImage*, 8(3):229–39, October 1998.

[81] Arthur K Liu, Jack W Belliveau, and Anders M Dale. Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulations. *Proceedings of the National Academy of Sciences*, 95(15):8945–50, July 1998.

[82] Matti S Hamalainen and J Sarvas. Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. *IEEE Transactions on Biomedical Engineering*, 36(2):165–71, February 1989.

[83] John C Mosher, Richard M Leahy, and P S Lewis. EEG and MEG: forward solutions for inverse methods. *IEEE Transactions on Biomedical Engineering*, 46(3):245–59, March 1999.

[84] John C Mosher, M E Spencer, Richard M Leahy, and P S Lewis. Error bounds for EEG and MEG dipole source localization. *Electroencephalography and Clinical Neurophysiology*, 86(5):303–21, May 1993.

[85] Kensuke Sekihara, Maneesh Sahani, and Srikantan S Nagarajan. Localization bias and spatial resolution of adaptive and non-adaptive spatial filters for MEG source reconstruction. *NeuroImage*, 25(4):1056–67, May 2005.

[86] Behtash Babadi, Gabriel Obregon-Henao, Camilo Lamus, Matti S Hämäläinen, Emery N Brown, and Patrick L Purdon. A subspace pursuit-based iterative greedy hierarchical solution to the neuromagnetic inverse problem. *NeuroImage*, 87:427–43, March 2014.

[87] Riitta Hari, Jari Karhu, Matti S. Hamalainen, Jukka Knuutila, Mikko Sams, and V Vilkman. Functional organization of the human first and second somatosensory cortices: a neuromagnetic study. *European Journal of Neuroscience*, 5:724–734, January 1993.

[88] Moshe Bar, Karim S Kassam, AS Ghuman, Jasmine Boshyan, AM Schmid, Anders M Dale, Matti S Hamalainen, K Marinkovic, DL Schacter, Bruce R Rosen, and Eric Halgren. Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103(2):449–454, January 2006.

[89] Mikko Sams, Reijo Aulanko, Matti S Hämäläinen, Riitta Hari, Olli V Lounasmaa, Sing-Teh Lu, and Juha Simola. Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, 127(1):141–5, June 1991.

[90] Daniel Baldauf and Robert Desimone. Neural mechanisms of object-based attention. *Science*, 344(6182):424–7, April 2014.

[91] Eran A Mukamel, Elvira Pirondini, Behtash Babadi, Kin Foon Kevin Wong, Eric T Pierce, P Grace Harrell, John L Walsh, Andres F Salazar-Gomez, Sydney S Cash, Emad N Eskandar, Veronica S Weiner, Emery N Brown, and Patrick L Purdon. A transition in brain state during propofol-induced unconsciousness. *The Journal of Neuroscience*, 34(3):839–45, January 2014.

[92] Sheraz Khan, Alexandre Gramfort, Nandita R Shetty, Manfred G Kitzbichler, Santosh Ganesan, Joseph M Moran, Su Mei Lee, John D E Gabrieli, Helen B Tager-Flusberg, Robert M Joseph, Martha R

Herbert, Matti S Hämäläinen, and Tal Kenet. Local and long-range functional connectivity is reduced in concert in autism spectrum disorders. *Proceedings of the National Academy of Sciences*, 110(8):3107–12, February 2013.

[93] Patrick L Purdon. *Multimodal neuroimaging with simultaneous electroencephalogram and high-field functional magnetic resonance imaging*. PhD thesis, Harvard University and Massachusetts Institute of Technology, 2005.

[94] Patrick L Purdon, Hernan Millan, Peter L Fuller, and Giorgio Bonmassar. An open-source hardware and software system for acquisition and real-time processing of electrophysiology during high-field MRI. *Journal of Neuroscience Methods*, 175(2):165–186, 2008.

[95] Frédéric Grouiller, Laurent Vercueil, Alexandre Krainik, Christoph Segebarth, Philippe Kahane, and Olivier David. A comparative study of different artefact removal algorithms for EEG signals acquired during functional MRI. *NeuroImage*, 38(1):124–37, October 2007.

[96] Girolamo Garreffa, Marta Bianciardi, Gisela E Hagberg, Emiliano Macaluso, Maria Grazia Marciani, Bruno Maraviglia, Manuel Abbafati, Marco Carnì, Ivo Bruni, and Luigi Bianchi. Simultaneous EEG-fMRI acquisition: how far is it from being a standardized technique? *Magnetic Resonance Imaging*, 22(10):1445–55, December 2004.

[97] Philip J Allen, Oliver Josephs, and Robert Turner. A method for removing imaging artifact from continuous EEG recorded during functional MRI. *NeuroImage*, 12(2):230–9, August 2000.

[98] Adriano B L Tort, Robert Komorowski, Howard Eichenbaum, and Nancy Kopell. Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *Journal of Neurophysiology*, 104(2):1195–210, August 2010.

[99] Kyle Q Lepage, Mark A Kramer, and Uri T Eden. Some sampling properties of common phase estimators. *Neural Computation*, 25(4):901–21, April 2013.

[100] Giorgio Bonmassar, Patrick L Purdon, Iiro P Jaaskelainen, Keith Chiappa, Victor Solo, Emery N Brown, and Jack W Beliveau. Motion and ballistocardiogram artifact removal for interleaved recording of EEG and EPs during MRI. *NeuroImage*, 16(4):1127–1141, August 2002.

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

[102] Rami K Niazy, C F Beckmann, Giandomenico D Iannetti, J M Brady, and Stephen M Smith. Removal of fMRI environment artifacts from EEG data using optimal basis sets. *NeuroImage*, 28(3):720–37, November 2005.

[103] Kyle Q Lepage, Christopher J Macdonald, Howard Eichenbaum, and Uri T Eden. The statistical analysis of partially confounded covariates important to neural spiking. *Journal of Neuroscience Methods*, 205(2):295–304, April 2012.

[104] Anders M Dale, Arthur K Liu, Bruce R Fischl, Randy L Buckner, Jack W Belliveau, Jeffrey D Lewine, Eric Halgren, and St Louis. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, 26:55–67, 2000.

[105] Fa-hsuan Lin, Thomas Witzel, Matti S Hämäläinen, Anders M Dale, Jack W Belliveau, and Steven M Stufflebeam. Spectral spatiotemporal imaging of cortical oscillations and interactions in the human brain. *Neuroimage*, 23(2):582–595, October 2004.

[106] Yohan Attal, Burkhard Maess, Angela Friederici, Olivier David, and Centre De. Head models and dynamic causal modeling of subcortical activity using magnetoencephalographic / electroencephalographic data. *Rev. Neurosci.*, 23(1):85–95, 2012.

[107] Eric Kandel, James Schwartz, and Thomas Jessell. *Principles of neural science*. McGraw-Hill Companies, fifth edition, 2000.

[108] Edward G Jones. The thalamic matrix and thalamocortical synchrony. *Trends in Neurosciences*, 24(10):595–601, October 2001.

[109] Edward G Jones. Thalamic circuitry and thalamocortical synchrony. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 357(1428):1659–73, December 2002.

[110] John Nolte. *The human brain: an introduction to its functional anatomy*. Mosby/Elsevier, sixth (illustrated) edition, 2009.

[111] Hal Blumenfeld. *Neuroanatomy through clinical cases*. Sinauer Associates, second edition, 2010.

[112] Yoshio C Okada, Lloyd Kaufman, and Samuel J Williamson. The hippocampal formation as a source of the slow endogenous potentials. *Electroencephalography and Clinical Neurophysiology*, 55(4):417–26, April 1983.

[113] Claudia D Tesche. Non-invasive imaging of neuronal population dynamics in human thalamus. *Brain Research*, 729(2):253–8, August 1996.

[114] Claudia D Tesche. Non-invasive detection of ongoing neuronal population activity in normal human hippocampus. *Brain Research*, 749(1):53–60, February 1997.

[115] Claudia D Tesche and Jari J Karhu. Anticipatory cerebellar responses during somatosensory omission in man. *Human Brain Mapping*, 9(3):119–42, March 2000.

[116] Yohan Attal, Manik Bhattacharjee, Jerome Yelnik, Benoit Cottereau, Julien Lefèvre, Yoshio Okada, Eric Bardinet, Marie Chupin, and Sylvain Baillet. Modeling and detecting deep brain activity with MEG & EEG. *Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007:4937–40, January 2007.

[117] Lauri Parkkonen, Nobuya Fujiki, and Jyrki P Makela. Sources of auditory brainstem responses revisited: contribution by magnetoencephalography. *Human Brain Mapping*, 30:1772–1782, April 2009.

[118] Christos Papadelis, Elisa Leonardelli, Martin Staudt, and Christoph Braun. Can magnetoencephalography track the afferent information flow along white matter thalamo-cortical fibers? *NeuroImage*, 60(2):1092–105, April 2012.

[119] Yohan Attal and Denis Schwartz. Assessment of subcortical source localization using deep brain activity imaging model with minimum norm operators: an MEG study. *PLoS One*, 8(3):e59856, January 2013.

[120] Hiroaki Ikeda, Leonard Leyba, Anton Bartolo, Yaozhi Wang, Yoshio C Okada, Domenico Restuccia, Catello Vollono, Daniela Virdis, Ivana Piero, Lucia Martucci, and Sergio Zanini. Synchronized spikes of thalamocortical axonal terminals and cortical neurons are detectable outside the pig brain with MEG. *Journal of Neurophysiology*, (87):626–630, January 2002.

[121] Maher A Quraan, Sandra N Moses, Yuwen Hung, Travis Mills, and Margot J Taylor. Detection and localization of hippocampal activity using beamformers with MEG: a detailed investigation using simulations and empirical data. *Human Brain Mapping*, 32(5):812–27, May 2011.

[122] Matti S Hamalainen and Risto J Ilmoniemi. Interpreting magnetic fields of the brain: minimum norm estimates. *Medical and Biological Engineering and Computing*, 32(1):35–42, January 1994.

[123] Peter D Gatehouse, Jennifer Keegan, Lindsey a Crowe, Sharmeen Masood, Raad H Mohiaddin, Karl-Friedrich Kreitner, and David N Firmin. Applications of phase-contrast flow and velocity imaging in cardiovascular MRI. *European Radiology*, 15(10):2172–84, October 2005.

[124] Gabor T Herman. *Fundamentals of Computerized Tomography*. Academic Press, second edition, 2010.

[125] Pavitra Krishnaswamy, Giorgio Bonmassar, Patrick L Purdon, and Emery N Brown. Reference-free harmonic regression technique to remove EEG-fMRI ballistocardiogram artifacts. *Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2013:5426–9, January 2013.

[126] Karsten Krakow, F G Woermann, M R Symms, Philip J Allen, Louis Lemieux, G J Barker, J S Duncan, and David R Fish. EEG-triggered functional MRI of interictal epileptiform activity in patients with partial seizures. *Brain*, 122:1679–1688, March 1999.

[127] Michael Czisch, Thomas C Wetter, Christian Kaufmann, Thomas Pollmächer, Florian Holsboer, and Dorothee P Auer. Altered processing of acoustic stimuli during sleep: reduced auditory activation and visual deactivation detected by a combined fMRI/EEG study. *NeuroImage*, 16(1):251–8, May 2002.

[128] Matthias Moosmann, Petra Ritter, Ina Krastel, Andrea Brink, Sebastian Thees, Felix Blankenburg, Birol Taskin, Hellmuth Obrig, and Arno Villringer. Correlates of alpha rhythm in functional magnetic resonance imaging and near infrared spectroscopy. *NeuroImage*, 20(1):145–58, September 2003.

[129] Vinod Menon, J M Ford, K O Lim, GH Glover, and A Pfefferbaum. Combined event-related fMRI and EEG evidence for temporal-parietal cortex activation during target detection. *NeuroReport*, 8(14):3029–3037, September 1997.

[130] Giorgio Bonmassar, D P Schwartz, Arthur K Liu, Kenneth K Kwong, Anders M Dale, and Jack W Belliveau. Spatiotemporal brain imaging of visual-evoked activity using interleaved EEG and fMRI recordings. *NeuroImage*, 13(6 Pt 1):1035–43, June 2001.

[131] Einat Liebenthal, Michael L Ellingson, Marianna V Spanaki, Thomas E Prieto, Kristina M Ropella, and Jeffrey R Binder. Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. *NeuroImage*, 19(4):1395–404, August 2003.

[132] Christoph Mulert, Lorenz Jäger, Robert Schmitt, Patrick Bussfeld, Oliver Pogarell, Hans-Jürgen Möller, Georg Juckel, and Ulrich Hegerl. Integration of fMRI and simultaneous EEG: towards a

comprehensive understanding of localization and time-course of brain activity in target detection. *NeuroImage*, 22(1):83–94, May 2004.

[133] Stefan Debener, Markus Ullsperger, Markus Siegel, Katja Fiehler, D Yves von Cramon, and Andreas K Engel. Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *The Journal of Neuroscience*, 25(50):11730–7, December 2005.

[134] R A Hill, K H Chiappa, Frank Huang-Hellinger, and Bruce G Jenkins. Hemodynamic and metabolic aspects of photosensitive epilepsy revealed by functional magnetic resonance imaging and magnetic resonance spectroscopy. *Epilepsia*, 40(7):912–920, July 1999.

[135] Frank Huang-Hellinger, Hans C. Breiter, Glen McCormack, Mark S. Cohen, Ken K. Kwong, Jeffrey P. Sutton, Robert L. Savoy, Robert M. Weisskoff, Timothly L. Davis, John R. Baker, Jack W. Belliveau, and Bruce R. Rosen. Simultaneous functional magnetic resonance imaging and electrophysiological recording. *Human Brain Mapping*, 3(1):13–23, October 1995.

[136] Stefan Debener, Karen J Mullinger, Rami K Niazy, and Richard W Bowtell. Properties of the ballistocardiogram artefact as revealed by EEG recordings at 1.5, 3 and 7 T static magnetic field strength. *International Journal for Psychophysiology*, 67(3):189–99, March 2008.

[137] Karen J Mullinger, Jade Havenhand, and Richard Bowtell. Identifying the sources of the pulse artefact in EEG recordings made inside an MR scanner. *NeuroImage*, 71:75–83, May 2013.

[138] J Sijbers, J Van Audekerke, M Verhoye, A Van der Linden, and D Van Dyck. Reduction of ECG and gradient related artifacts in simultaneously recorded human EEG/MRI data. *Magnetic Resonance Imaging*, 18(7):881–6, September 2000.

[139] Michael L Ellingson, Einat Liebenthal, M V Spanaki, T E Prieto, J R Binder, and Kristina M Ropella. Ballistocardiogram artifact reduction in the simultaneous acquisition of auditory ERPs and fMRI. *NeuroImage*, 22(4):1534–42, August 2004.

[140] Francesco Musso, Jürgen Brinkmeyer, Daniel Ecker, Markus K London, Giesela Thieme, Tracy Warbrick, Hans-Jörg Wittsack, Andreas Saleh, Wolfgang Greb, Peter de Boer, and Georg Winterer. Ketamine effects on brain function - simultaneous fMRI/EEG during a visual oddball task. *NeuroImage*, 58(2):508–525, June 2011.

[141] Kyung Hwan Kim, Hyo Woon Yoon, and Hyun Wook Park. Improved ballistocardiac artifact removal from the electroencephalogram recorded in fMRI. *Journal of Neuroscience Methods*, 135(1-2):193–203, May 2004.

[142] Wakako Nakamura, Kimitaka Anami, Takeyuki Mori, Osamu Saitoh, Andrzej Cichocki, and Shun-ichi Amari. Removal of ballistocardiogram artifacts from simultaneously recorded EEG and fMRI data using independent component analysis. *IEEE Transactions on Biomedical Engineering*, 53(7):1294–308, July 2006.

[143] Dante Mantini, M G Perrucci, S Cugini, A Ferretti, G L Romani, and C Del Gratta. Complete artifact removal for EEG recorded during continuous fMRI using independent component analysis. *NeuroImage*, 34(2):598–607, January 2007.

[144] Stefan Debener, Alexander Strobel, Bettina Sorger, Judith Peters, Cornelia Kranczioch, Andreas K Engel, and Rainer Goebel. Improved quality of auditory event-related potentials recorded simultaneously with 3T fMRI: removal of the ballistocardiogram artefact. *NeuroImage*, 34(2):587–97, January 2007.

[145] Xiaohong Wan, Kazuki Iwata, Jorge Riera, Torh Ozaki, Masaharu Kitamura, and Ryuta Kawashima. Artifact reduction for EEG/fMRI recording: nonlinear reduction of ballistocardiogram artifacts. *Clinical Neurophysiology*, 117(3):668–80, March 2006.

[146] Katrien Vanderperren, Jennifer Ramautar, Nikolai Novitski, and Maarten De Vos. Ballistocardiogram artifacts in simultaneous EEG- fMRI acquisitions. *International Journal of Bioelectromagnetism*, 9(3):146–150, June 2007.

[147] Justin L Vincent, Linda J Larson-Prior, John M Zempel, and Abraham Z Snyder. Moving GLM ballistocardiogram artifact reduction for EEG acquired simultaneously with fMRI. *Clinical Neurophysiology*, 118(5):981–98, May 2007.

[148] Henry Gray. *Anatomy of the human body*. Bartleby.Com, Philadelphia, twentieth edition, 2000.

[149] Matthew Brandon Maas and Joseph Safdieh. Ischemic stroke: pathophysiology and principles of localization. *Neurology*, 13(Pt. 1):1–16, Board Review Manual 2009.

[150] Patrick L Purdon, Eric T Pierce, Eran A Mukamel, Michael J Prerau, Kin Foon Wong, Andres F Salazar-Gomez, Casie L Tavares-Stoeckel, Kathleen Habeeb, Rebecca Merhar, and Emery N Brown. Electroencephalogram signatures of loss and recovery of consciousness from propofol. *Proceedings of the National Academy of Sciences*, 110(12):E1142–E1151, 2013.

[151] Craig F Ansley and William E Wecker. The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association*, 78(381):81–89, March 1983.

[152] Emery N Brown and Christopher H Schmid. Application of the Kalman filter to computational problems in statistics. *Methods in Enzymology*, 240:171–181, 1994.

[153] Emery N Brown, Victor Solo, Yong Choe, and Zhenhua Zhang. Measuring period of human biological clock: infill asymptotic analysis of harmonic regression parameter estimates. *Methods in Enzymology*, 383(1989):382–405, January 2004.

[154] Wasim Q Malik, James Schummers, Mriganka Sur, and Emery N Brown. Denoising two-photon calcium imaging data. *PLoS One*, 6(6):e20490, January 2011.

[155] Robert Tibshirani and Trevor Hastie. Local likelihood estimation. *Journal of the American Statistical Association*, 82(398):559–567, June 1987.

[156] Corrado Corradi. A note on the computation of maximum likelihood estimates in linear regression models with autocorrelated errors. *Journal of Econometrics*, 11:303–317, January 1979.

[157] Giandomenico D Iannetti, Rami K Niazy, R G Wise, P Jezzard, J C W Brooks, L Zambreanu, W Vennart, P M Matthews, and Irene Tracey. Simultaneous recording of laser-evoked brain potentials and continuous, high-field functional magnetic resonance imaging in humans. *NeuroImage*, 28(3):708–19, November 2005.

[158] Arnaud Delorme and Scott Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, 134:9–21, October 2004.

[159] Francesco Di Russo, Martin I Sereno, Sabrina Pitzalis, and Steven A Hillyard. Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, 15:95–111, September 2001.

[160] R Beisteiner, S Robinson, M Wurnig, M Hilbert, K Merksa, J Rath, I Höllinger, N Klinger, Ch Marosi, S Trattnig, and A Geissler. Clinical fMRI: evidence for a 7T benefit over 3T. *NeuroImage*, 57(3):1015–21, August 2011.

[161] Federico De Martino, Fabrizio Esposito, Pierre-Francois van de Moortele, Noam Harel, Elia Formisano, Rainer Goebel, Kamil Ugurbil, and Essa Yacoub. Whole brain high-resolution functional imaging at ultra high magnetic fields: an application to the analysis of resting state networks. *NeuroImage*, 57(3):1031–44, August 2011.

[162] Irene Neuner, Tracy Warbrick, Jorge Arrubla, Jörg Felder, Avdo Celik, Martina Reske, Franks Boers, and N Jon Shah. EEG acquisition in ultra-high static magnetic fields up to 9.4 T. *NeuroImage*, 68:214–20, March 2013.

[163] Jack R Foucher, Hélène Otzenberger, and Daniel Gounot. The BOLD response and the gamma oscillations respond differently than evoked potentials: an interleaved EEG-fMRI study. *BMC Neuroscience*, 4(22):1–11, September 2003.

[164] Jean-Philippe Lachaux, Pierre Fonlupt, Philippe Kahane, Lorella Minotti, Dominique Hoffmann, Olivier Bertrand, and Monica Baciu. Relationship between task-related gamma oscillations and BOLD signal: new insights from combined fMRI and intracranial EEG. *Human Brain Mapping*, 28(12):1368–75, December 2007.

[165] Anders M Dale and Eric Halgren. Spatiotemporal mapping of brain activity by integration of multiple imaging modalities. *Current Opinion in Neurobiology*, 11(2):202–8, April 2001.

[166] Fa-Hsuan Lin, Thomas Witzel, Seppo P Ahlfors, Steven M Stufflebeam, Jack W Belliveau, and Matti S Hämäläinen. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage*, 31(1):160–71, May 2006.

[167] John C Mosher, P S Lewis, and Richard M Leahy. Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Transactions on Biomedical Engineering*, 39(6):541–57, June 1992.

[168] Kimmo Uutela, Matti S Hämäläinen, and Riitta Salmelin. Global optimization in the localization of neuromagnetic sources. *IEEE Transactions on Biomedical Engineering*, 45(6):716–23, June 1998.

[169] Masa-aki Sato, Taku Yoshioka, Shigeki Kajihara, Keisuke Toyama, Naokazu Goda, Kenji Doya, and Mitsuo Kawato. Hierarchical Bayesian estimation for MEG inverse problem. *NeuroImage*, 23(3):806–26, November 2004.

[170] David Wipf and Srikantan Nagarajan. A unified Bayesian framework for MEG/EEG source imaging. *NeuroImage*, 44(3):947–66, February 2009.

[171] Andreas Galka, Okito Yamashita, Tohru Ozaki, Rolando Biscay, and Pedro Valdés-Sosa. A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering. *NeuroImage*, 23(2):435–53, October 2004.

[172] Camilo Lamus, Matti S Hamalainen, Simona Temereanca, Emery Brown, and Patrick Purdon. Estimation and identification in spatiotemporal dynamic distributed models for the MEG inverse problem. *NeuroImage*, 63(2):894–909, March 2012.

[173] Muthuraman Muthuraman, Helge Hellriegel, Nienke Hoogenboom, Abdul Rauf Anwar, Kidist Gebremariam Mideksa, Holger Krause, Alfons Schnitzler, Günther Deuschl, and Jan Raethjen. Beamformer source analysis and connectivity on concurrent EEG and MEG data during voluntary movements. *PLoS One*, 9(3):e91441, January 2014.

[174] Panagiotis G Simos, L F Basile, and Andrew C Papanicolaou. Source localization of the N400 response in a sentence-reading paradigm using evoked magnetic fields and magnetic resonance imaging. *Brain Research*, 762(1-2):29–39, July 1997.

[175] Ribary Urs, A A Ionnides, Krish D Singh, Robert Hasson, J Stuart Bolton, Fred Lado, Alon Mogilner, and Rodolfo Llinas. Magnetic field tomography of coherent thalamocortical 40-Hz oscillations in humans. *Proceedings of the National Academy of Sciences*, 88:11037–11041, December 1991.

[176] Liza Gross. From structure to function: mapping the connection matrix of the human brain. *PLoS Biology*, 6(7):e164, July 2008.

[177] Karim Jerbi, Jean-Philippe Lachaux, Karim N'Diaye, Dimitrios Pantazis, Richard M Leahy, Line Garnero, and Sylvain Baillet. Coherent neural representation of hand speed in humans revealed by MEG imaging. *Proceedings of the National Academy of Sciences*, 104(18):7676–81, May 2007.

[178] Anders M Dale, Bruce Fischl, and Martin I Sereno. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, 9:179–194, January 1999.

[179] Bruce Fischl, Martin I Sereno, and Anders M Dale. Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9:195–207, January 1999.

[180] Bruce Fischl, David H Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre van der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, Albert Montillo, Nikos Makris, Bruce Rosen, and Anders M Dale. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–55, January 2002.

[181] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Lauri Parkkonen, and Matti S Hämäläinen. MNE software for processing MEG and EEG data. *NeuroImage*, 86:446–60, February 2014.

[182] Tulaya Limpiti, Barry D Van Veen, and Ronald T Wakai. Cortical patch basis model for spatially extended neural activity. *IEEE Transactions on Biomedical Engineering*, 53(9):1740–54, September 2006.

[183] Ake Bjorck and Gene H Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, July 1973.

[184] Andrew V Knyazev and Merico E Argentati. Principal angles between subspaces in an A-based scalar product: algorithms and perturbation estimates. *SIAM Journal on Scientific Computing*, 23(6):2008–2040, February 2002.

[185] Jerome H Friedman and John Tukey. A projection pursuit algorithm for exploratory data analysis. *Stanford Linear Accelerator Center Publication*, 1312:1–39, September 1973.

[186] Peter Huber. Projection pursuit. *The Annals of Statistics*, 13(2):435–475, January 1985.

[187] Deanna Needell and Joel A Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, May 2009.

[188] Wei Dai and Olgica Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Transactions on Information Theory*, 55(5):2230–2249, May 2009.

[189] Matti S Hamalainen. MNE software user's guide, version 2.7.3. *Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA*, 2010.

[190] Alexandre Gramfort, Daniel Strohmeier, Jens Haueisen, Matti S Hamalainen, and Matthieu Kowalski. Functional brain imaging with M/EEG using structured sparsity in time-frequency dictionaries. *Proceedings: Conference on Information Processing in Medical Imaging*, 22:600–11, January 2011.

[191] Susan Pockett. Anesthesia and the electrophysiology of auditory consciousness. *Consciousness and Cognition*, 61:45–61, 1999.

[192] Jyrki P Makela, Matti S Hamalainen, Riitta Hari, and Linda McEvoy. Whole-head mapping of middle-latency auditory evoked magnetic fields. *Electroencephalography and Clinical Neurophysiology*, 92(5):414–21, September 1994.

[193] Irina Gorodnitsky, John George, and Bhaskar Rao. Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. *Electroencephalography and Clinical Neurophysiology*, 95:231–251, May 1995.

[194] Mayrim Vega-Hernandez, Eduardo Martinez-Montes, Jose Sanchez-Bornot, Agustin Lage-Castellanos, and Pedro Valdes-Sosa. Penalized least squares methods for solving the EEG inverse problem. *Statistica Sinica*, 18:1535–1551, 2008.

[195] Alexandre Gramfort, Matthieu Kowalski, and Matti S Hämäläinen. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Physics in Medicine and Biology*, 57(7):1937–61, April 2012.

[196] Alexandre Gramfort, Daniel Strohmeier, Jens Haueisen, Matti S Hämäläinen, and Matthieu Kowalski. Time-frequency mixed-norm estimates: sparse M/EEG imaging with non-stationary source activations. *NeuroImage*, 70:410–422, January 2013.

[197] Wanmei Ou, Matti S Hämäläinen, and Polina Golland. A distributed spatio-temporal EEG/MEG inverse solver. *NeuroImage*, 44(3):932–46, February 2009.

[198] Piotr J Durka, Artur Matysiak, Eduardo Martínez Montes, Pedro Valdés Sosa, and Katarzyna J Blinowska. Multichannel matching pursuit and EEG inverse solutions. *Journal of Neuroscience Methods*, 148(1):49–59, October 2005.

[199] Shinung Ching, Aylin Cimenser, Patrick L Purdon, Emery N Brown, and Nancy J Kopell. Thalamocortical model for a propofol-induced alpha-rhythm associated with loss of consciousness. *Proceedings of the National Academy of Sciences*, 107(52):22665–70, December 2010.

[200] Michael D Fox, Abraham Z Snyder, Justin L Vincent, Maurizio Corbetta, David C Van Essen, and Marcus E Raichle. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, 102(27):9673–8, July 2005.

[201] Jessica S Damoiseaux, Serge A R B Rombouts, Frederik Barkhof, P Scheltens, Cornelis J Stam, Stephen M Smith, and Christian F Beckmann. Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences*, 103(37):13848–13853, September 2006.

[202] Helga A. Harsay, Michael X Cohen, Nikolaas N Oosterhof, Birte U Forstmann, Rogier B Mars, and K. Richard Ridderinkhof. Functional connectivity of the striatum links motivation to action control in humans. *Journal of Neuroscience*, 31(29):10701–10711, July 2011.

[203] Markus Siegel, Tobias H Donner, and Andreas K Engel. Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews Neuroscience*, 13(2):121–34, February 2012.

[204] Andrew J Watrous, Nitin Tandon, Chris R Conner, Thomas Pieters, and Arne D Ekstrom. Frequency-specific network connectivity increases underlie accurate spatiotemporal memory retrieval. *Nature Neuroscience*, 16(3):349–56, March 2013.

[205] Emery N Brown, Patrick L Purdon, and Christa J Van Dort. General anesthesia and altered states of arousal: a systems neuroscience analysis. *Annual Review of Neuroscience*, 34:601–628, April 2011.

[206] Nicholas P Franks. General anaesthesia: from molecular targets to neuronal pathways of sleep and arousal. *Nature Reviews Neuroscience*, 9(5):370–86, May 2008.

[207] Anne B Young and Dorothy Chu. Distribution of GABAa and GABAb receptors in mammalian brain: potential targets for drug development. *Drug Development Research*, 21:161–167, July 1990.

[208] Tim G Hales and Jeremy J Lambert. The actions of propofol on inhibitory amino acid receptors of bovine adrenomedullary chromaffin cells and rodent central neurons. *British Journal of Pharmacology*, 104(3):619–28, November 1991.

[209] Manami Hara, Yoshihisa Kai, and Yoshimi Ikemoto. Propofol activates GABAa receptor-chloride ionophore complex in dissociated hippocampal pyramidal neurons of the rat. *Anesthesiology*, 79(4):781–788, October 1993.

[210] Lee A Kearse, Emery N Brown, and Kathleen McPeck. Somatosensory evoked potentials sensitivity relative to electroencephalography for cerebral ischemia during carotid endarterectomy. *Stroke*, 23(4):498–505, April 1992.

[211] F C Forrest, M A Tooley, P R Saunders, and C Prys-Roberts. Propofol infusion and the suppression of consciousness: the EEG and dose requirements. *British Journal of Anaesthesia*, 72(1):35–41, January 1994.

[212] R V Reddy, S S Moorthy, S F Dierdorf, R D Deitch, and L Link. Excitatory effects and electroencephalographic correlation of etomidate, thiopental, methohexital, and propofol. *Anesthesia and Analgesia*, 77(5):1008–11, November 1993.

[213] Michael Murphy, Marie-Aurélie Bruno, Brady A Riedner, Pierre Boveroux, Quentin Noirhomme, Eric C Landsness, Jean-Francois Brichant, Christophe Phillips, Marcello Massimini, Steven Laureys, Giulio Tononi, and Mélanie Boly. Propofol anesthesia and sleep: a high-density EEG study. *Sleep*, 34(3):283–291A, March 2011.

[214] Laura D Lewis, Veronica S Weiner, Eran A Mukamel, Jacob A Donoghue, Emad N Eskandar, Joseph R Madsen, William S Anderson, Leigh R Hochberg, Sydney S Cash, Emery N Brown, and Patrick L Purdon. Rapid fragmentation of neuronal networks at the onset of propofol-induced unconsciousness. *Proceedings of the National Academy of Sciences*, 109(49):E3377–86, December 2012.

[215] Laverne D Gugino, Robert J Chabot, Leslie S Prichep, E R John, V Formanek, and Linda Aglio. Quantitative EEG changes associated with loss and return of consciousness in healthy adult volunteers anaesthetized with propofol or sevoflurane. *British Journal of Anaesthesia*, 87(3):421–428, May 2001.

[216] E R John, Leslie S Prichep, W Kox, Peter Valdés-Sosa, J Bosch-Bayard, E Aubert, M Tom, F di Michele, and Laverne D Gugino. Invariant reversible QEEG effects of anesthetics. *Consciousness and Cognition*, 10(2):165–83, June 2001.

[217] Vladimir Feshchenko, Robert Veselis, and Ruth Reinsel. Propofol-induced alpha rhythm. *Neuropsychobiology*, 50(3):257–66, January 2004.

[218] Aylin Cimenser, Patrick L Purdon, Eric T Pierce, John L Walsh, Andres F Salazar-Gomez, Priscilla G Harrell, Casie Tavares-Stoeckel, Kathleen Habeeb, and Emery N Brown. Tracking brain states under general anesthesia by using global coherence analysis. *Proceedings of the National Academy of Sciences*, 108(21):8832–8837, May 2011.

[219] Vau Jäntti, A Yli-Hankala, G A Baer, and T Porkkala. Slow potentials of EEG burst suppression pattern during anaesthesia. *Acta Anaesthesiologica Scandinavica*, 37(1):121–3, January 1993.

[220] William P Akrawi, John C Drummond, Cor J Kalkman, and Piyush M Patel. A comparison of the electrophysiologic characteristics of EEG burst-suppression as produced by isoflurane, thiopental, etomidate, and propofol. *Journal of Neurosurgical Anesthesiology*, 8(1):40–6, January 1996.

[221] Gernot G. Supp, Markus Siegel, Joerg F. Hipp, and Andreas K. Engel. Cortical hypersynchrony predicts breakdown of sensory processing during loss of consciousness. *Current Biology*, 21(23):1–6, November 2011.

[222] Jonathan D Breshears, Jarod L Roland, Mohit Sharma, Charles M Gaona, Zachary V Freudenburg, Rene Tempelhoff, Michael S Avidan, and Eric C Leuthardt. Stable and dynamic cortical electrophysiology of induction and emergence with propofol anesthesia. *Proceedings of the National Academy of Sciences*, 107(49):21170–5, December 2010.

[223] Nicholas P Franks and Anna Y Zecharia. Sleep and general anesthesia. *Canadian Journal of Anaesthesia*, 58(2):139–48, February 2011.

[224] Michael Alkire, Richard Haier, Steven Barker, Nitin Shah, Joseph Wu, and Kao James. Cerebral metabolism during propofol anesthesia in humans studied with positron emission tomography. *Anesthesiology*, 82(2):393–403, January 1995.

[225] Pierre Fiset, Tomas Paus, Thierry Daloze, Gilles Plourde, Pascal Meuret, Vincent Bonhomme, Nadine Hajj-Ali, Steven B Backman, and Alan C Evans. Brain mechanisms of propofol-induced loss of consciousness in humans: a positron emission tomographic study. *The Journal of Neuroscience*, 19(13):5506–13, July 1999.

[226] Michael T Alkire, Richard J Haier, and James H Fallon. Toward a unified theory of narcosis: brain imaging evidence for a thalamocortical switch as the neurophysiologic basis of anesthetic-induced unconsciousness. *Consciousness and Cognition*, 9(3):370–86, September 2000.

[227] Jaakko W Langsjo, Michael T Alkire, Kimmo Kaskinoro, Hiroki Hayama, Anu Maksimow, Kaike K Kaisti, Sargo Aalto, Riku Aantaa, Satu K Jaaskelainen, Antti Revonsuo, and Harry Scheinin. Returning from oblivion: imaging the neural core of consciousness. *Journal of Neuroscience*, 32(14):4935–4943, April 2012.

[228] Ithabi Ganter, Pieter Guldenmund, Francisco Gomez, Audrey Venhadenhuyse, Pierre Boverous, Steven Laureys, and Andrea Soddu. Propofol induced unconsciousness: fMRI total neuronal activity and resting state networks. Frontiers Abstract, 2007.

[229] Pierre Boveroux, Audrey Vanhaudenhuyse, Marie-Aurelie Bruno, Quentin Noirhomme, Severine Lauwick, Andre Luxen, Christian Degueldre, Alain Plenevaux, Caroline Schnakers, Christophe Phillips, Jean-Francois Brichant, Vincent Bonhomme, Pierre Maquet, Michael D Greicius, Steven Laureys, and Melanie Boly. Breakdown of within- and between-network resting state connectivity during propofol-induced loss of consciousness. *Anesthesiology*, 113(5):1038–53, November 2010.

[230] Jessica Schrouff, Vincent Perlbarg, Mélanie Boly, Guillaume Marrelec, Pierre Boveroux, Audrey Vanhaudenhuyse, Marie-Aurélie Bruno, Steven Laureys, Christophe Phillips, Mélanie Pélégrini-Issac, Pierre Maquet, and Habib Benali. Brain functional integration decreases during propofol-induced loss of consciousness. *NeuroImage*, 57(1):198–205, July 2011.

[231] Manuel S Schröter, Victor I Spoormaker, Anna Schorer, Afra Wohlschläger, Michael Czisch, Eberhard F Kochs, Claus Zimmer, Bernhard Hemmer, Gerhard Schneider, Denis Jordan, and Rüdiger Ilg. Spatiotemporal reconfiguration of large-scale brain functional networks during propofol-induced loss of consciousness. *The Journal of Neuroscience*, 32(37):12832–40, September 2012.

[232] Róisín Ní Mhuircheartaigh, Catherine Warnaby, Richard Rogers, Saad Jbabdi, and Irene Tracey. Slow-wave activity saturation and thalamocortical isolation during propofol anesthesia in humans. *Science Translational Medicine*, 5(208):208ra148, October 2013.

[233] George A Mashour and Michael T Alkire. Consciousness, anesthesia, and the thalamocortical system. *Anesthesiology*, 118(1):13–15, January 2013.

[234] Partha Mitra and Hemant Bokil. *Observed brain dynamics*. Oxford University Press, New York, first edition, 2008.

[235] Aaron L Sampson, Behtash Babadi, Michael J Prerau, Eran A Mukamel, Emery N Brown, and Patrick L Purdon. Beamforming approach to phase-amplitude modulation analysis of multi-channel EEG. *Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012:6731–4, January 2012.

[236] Mark W Woolrich, Timothy E J Behrens, Christian F Beckmann, Mark Jenkinson, and Stephen M Smith. Multilevel linear modelling for fMRI group analysis using Bayesian inference. *NeuroImage*, 21(4):1732–47, April 2004.

[237] Michael Lagerkranser, Kristina Stange, and Alf Sollevi. Effects of propofol on cerebral blood flow, metabolism and cerebral autoregulation in the anesthetized pig. *Journal of Neurosurgical Anesthesiology*, 9(2):188–193, April 1997.

[238] M Enlund, J Andersson, P Hartvig, J Valtysson, and L Wiklund. Cerebral normoxia in the rhesus monkey during isoflurane- or propofol-induced hypotension and hypocapnia, despite disparate blood-flow patterns: a positron emission tomography study. *Acta Anaesthesiologica Scandinavica*, 41(8):1002–10, September 1997.

[239] Eroglu Basar, Martin Schürmann, Canan Basar-Eroglu, and Sirel Karakas. Alpha oscillations in brain functioning: an integrative theory. *International journal of Psychophysiology*, 26(1-3):5–29, June 1997.

[240] Anthony Angel. The G. L. Brown lecture: adventures in anaesthesia. *Experimental Physiology*, 76(1):1–38, January 1991.

[241] Michael Avidan, Eric Jacobsohn, David Glick, Beth Burnside, Lini Zhang, Alex Villafranca, Leah Karl, Saima Kamal, Brian Torres, Michael O Connor, Alex S Evers, Stephen Gradwohl, Nan Lin, Ben J Palanca, and George Mashour. Prevention of intraoperative awareness in a high-risk surgical population. *New England Journal of Medicine*, 365(7):591–600, 2011.

[242] Peter S Sebel, T Andrew Bowdle, Mohamed M Ghoneim, Ira J Rampil, Roger E Padilla, Tong Joo Gan, and Karen B Domino. The incidence of awareness during anesthesia: a multicenter United States study. *Anesthesia and Analgesia*, 99(3):833–9, September 2004.

[243] L Krenk and L S Rasmussen. Postoperative delirium and postoperative cognitive dysfunction in the elderly - what are the differences? *Minerva Anestesiologica*, 77(7):742–9, July 2011.

[244] Ryan Vaden, Nathan Hutcheson, Lesley McCollum, Johnny Kentros, and Kristina Visscher. Older adults, unlike younger adults, do not modulate alpha power to suppress irrelevant information. *Neuroimage*, 63(3):1127–1133, November 2012.

[245] Emery N Brown and Patrick L Purdon. The aging brain and anesthesia. *Current Opinion in Anaesthesiology*, 26(4):414–419, August 2013.

[246] Stuart W Hughes and Vincenzo Crunelli. Thalamic mechanisms of EEG alpha rhythms and their pathological implications. *The Neuroscientist*, 11(4):357–72, August 2005.

[247] Vincenzo Crunelli and Stuart W Hughes. The slow rhythm of non-REM sleep: a dialogue between three cardinal oscillators. *Nature Neuroscience*, 13(1):9–17, January 2010.

[248] Kimitaka Anami, Takeyuki Mori, Fumiko Tanaka, Yusuke Kawagoe, Jun Okamoto, Masaru Yarita, Takashi Ohnishi, Masato Yumoto, Hiroshi Matsuda, and Osamu Saitoh. Stepping stone sampling for retrieving artifact-free electroencephalogram during functional magnetic resonance imaging. *NeuroImage*, 19(2):281–295, June 2003.

[249] Christos E Vasios, Leonardo M Angelone, Patrick L Purdon, Jyrki Ahveninen, Jack W Belliveau, and Giorgio Bonmassar. EEG/(f)MRI measurements at 7 Tesla using a new EEG cap ("InkCap"). *NeuroImage*, 33(4):1082–92, December 2006.

[250] Van J Wedeen, Douglas L Rosene, Ruopeng Wang, Guangping Dai, Farzad Mortazavi, Patric Hagmann, Jon H Kaas, and Wen-Yih I Tseng. The geometric structure of the brain fiber pathways. *Science*, 335(6076):1628–34, March 2012.

[251] Mircea Steriade, Angel Nuiiez, and Florin Amzica. A novel slow oscillation of neocortical neurons in vivo: depolarizing and hyperpolarizing components. *The Journal of Neuroscience*, 13(8):3252–65, August 1993.

[252] Mircea Steriade. Sleep oscillations and their blockage by activating systems. *Journal of Psychiatry & Neuroscience*, 19(5):354–8, November 1994.

[253] Igor Timofeev, Francois Grenier, and Mircea Steriade. Disfacilitation and active inhibition in the neocortex during the natural sleep-wake cycle: an intracellular study. *Proceedings of the National Academy of Sciences*, 98(4):1924–9, February 2001.

[254] Mircea Steriade, Marian Tsanov, Ehsan Chah, Nick Wright, Seralynne D Vann, Richard Reilly, T Jonathan, John P Aggleton, Shane M O Mara, and Xiao-jing Wang. Impact of network activities on neuronal properties in corticothalamic systems. *Journal of Neurophysiology*, 86:1–39, March 2012.