Characterizing response to patterned electrical stimulation to improve neuromodulation therapies

Samantha H. Sun

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2024

Reading Committee:
Rajesh Rao, Chair
Eric Chudler
Amy Orsborn

Program Authorized to Offer Degree:
Bioengineering

University of Washington

Abstract

Characterizing response to patterned electrical stimulation to improve neuromodulation therapies

Samantha H. Sun

Co-chairs of the Supervisory Committee:

Jeffrey Ojemann
Department of Neurological Surgery

Rajesh Rao
School of Computer Science and Engineering

Neuromodulation therapies, such as deep brain stimulation (DBS), aim to modulate dysfunctional neural circuitry in patients with treatment-resistant neurological disorders for symptom alleviation. DBS applications extend from reducing motor impairments in Parkinson's Disease, essential tremor, and dystonia, to improving mood-related symptoms in major depressive disorders, obsessive compulsive disorder, and bipolar disorder. Adaptive deep brain stimulation (aDBS) has become a critical advancement in neuromodulation, where stimulation therapy is adjusted based on current patient needs, and it has led to improved patient outcomes, reduced side-effects, and increased device battery life. Leveraging the unique opportunity of recording from invasive electroencephalography (iEEG) in humans, this thesis contributes to three distinct components of adaptive DBS: biomarker detection for brain-state estimation, characterizing stimulation response, and identifying candidate stimulation sites for effective neuromodulation.

We first motivate the importance of brain-state estimation for adaptive neuromodulation and demonstrate two examples: an unsupervised sleep/wake classifier and a subject-specific pain state classifier. We first build an unsupervised sleep/wake classifier using a Hidden Semi-Markov Model on several-hour long datasets. We showed that our model outperforms other

unsupervised models and discuss how unsupervised models could be leveraged in situations where data are plenty, but labels are sparse. We also build a subject-specific pain state classifier and identify unique neural features and models that predict subject-reported pain levels. We identify that models performed better in some subjects and discuss how differences in subject-reported pain may contribute to model performance.

Next, we characterize stimulation response by performing patterned electrical stimulation by varying stimulation amplitude and frequency and measuring the responses to stimulation at many recording channels. We identified in what channels stimulation response could be detected and to what extent different stimulation patterns were distinguishable. We found that most subjects had few, select channels that are capable of distinguishing between different stimulation patterns, and that some but not all stimulation patterns are separable. These results contribute to our foundational understanding of neuromodulation, suggesting directions for future work to continue towards building a general model for stimulation response.

Finally, we present results establishing that single-pulse stimulation response can be used to predict a channel's sensitivity to patterned stimulation, providing an efficient method to identify channels responsive to stimulation. We then leverage this idea and demonstrate an experimental paradigm to create stimulation response maps across many stimulation channels. Given a target site for neuromodulation, these stimulation response maps help identify candidate stimulation sites, which could be a helpful tool in identifying patient-specific sites for neuromodulation.

Acknowledgments

Many communities have provided immeasurable encouragement, kindness, and support throughout my academic journey.

I would first like to thank my co-advisers, Rajesh Rao and Jeffrey Ojemann, for providing the foundational support for our lab, the rare opportunities for working in the human intracranial space, and for their guidance in both my scientific and career pursuits. I'd like to thank my committee members present and past— Amy Orsborn, Sara Goering, Eric Chudler, Azadeh Yazdan, and Bingni Brunton—for providing invaluable feedback and encouragement along the way. I also want to thank Jeffrey Herron for stepping into a primary mentorship role for many students, including myself, and for generously giving the space and time for important conversations regarding research and beyond.

Having the opportunity to work in a highly collaborative space, I was lucky to have met and worked with many lab members between the Neural Systems Lab and the GRIDlab—Zoe Steine-Hanson, Ellie Strandquist, Raphael Bechtold, Hanbin Cho, Jesse Miles, and Sophia Lowe-Hones. I wanted to thank Preston Jiang, Timmy Pham, Lila Levinson, and Courtnie Paschall, as working with you all on research or data collection has been a highlight of my time here. I also want to thank GRIDlab alums Nile Wilson, Jenny Cronin, David Caldwell, James Wu, and Tim Brown, as they have been important mentors and role models for me.

I'd like to thank the patients that I had the pleasure of working with, as they are the reason this research is possible. I'd like to thank the entire clinical team, including neurosurgeons Andrew Ko, Jason Hauptman, and Ben Grannan, and the EEG monitoring team including Shahin Hakimian and Jeffrey Tsai.

Thank you to my friends and family, near and far, for believing in me, being patient with me, and reminding me to make space for joy, especially in that final year.

Finally, I'd like to thank my first undergraduate research advisor, Christine MacDonald. Your kind but firm values-based mentorship still sits with me to this day. Thank you for advocating for me, cheering me on, and showing me the confidence to be a woman in STEM.

Table of contents

List of Figures

List of Tables

1. Introduction

1.1 Electrical stimulation as therapeutic neuromodulation

Electrical stimulation is a powerful tool for targeted neurorehabilitation, with demonstrated utility in reducing tremor symptoms using deep brain stimulation (DBS) for Parkinson's Disease and essential tremor[1], interrupting seizure onset using responsive stimulation for epilepsy[2], and recovering vital motor functions after spinal cord injury using spinal cord stimulation[3].

A growing area of interest for electrical stimulation research is adaptive or responsive stimulation, where stimulation is adjusted in real-time, informed by a neural biomarker, to reduce stimulation-induced side effects[4,5] and/or improve stimulation outcomes[6–8]. While advancements have been made in decoding these neural biomarkers in promising applications such as for depression[9] and obsessive-compulsive disorder[10], in addition to advancements in adaptive decoding models that change as the brain learns[11], less is known about the mechanisms of stimulation—the encoding process—and how stimulation interacts with underlying neural activity. As such, there is no existing model that informs how to adjust electrical stimulation in real-time to reach a desired neural or behavioral outcome.

1.2 Challenges of stimulation work

Lack of population-level models for predicting electrical stimulation response
Development of adaptive electrical stimulation is limited by our lack of understanding of how electrical stimulation is processed in the brain. Electrical stimulation activates populations of neurons, up to hundreds of thousands near the site of stimulation, and response to stimulation is inextricably linked to the neuroanatomy surrounding the site of stimulation[12]. Therefore, generalizability of stimulation response is questionable across brain regions and even more so across individuals. In addition to static properties of structural anatomy, the dynamic nature of the brain and its ever-changing states can elicit different responses to electrical stimulation, even when the stimulation input remains identical[13]. The dependence of stimulation on its static location and the influence of ongoing temporal dynamics makes creating a general model of electrical stimulation at a population level difficult to attain. However, to build adaptive stimulation methods, there is a need for a model of electrical

stimulation and how it interacts with underlying activity to produce our recorded neural response.

Existing models of electrical stimulation are limited in scale and are unable to inform adaptive stimulation in population-level effects. Biophysical neuron models, such as the Hodgkin-Huxley model[14] and integrate and fire neuron model[15], can explain firing rate changes in single neurons when given an injected current input. Mid-scale plasticity models, where learning rules such as Hebbian plasticity and spike-timing dependent plasticity[16] can be incorporated with these single neuron models to reproduce stimulation experiments and explain how stimulation input can alter synaptic weight between neurons[17]. However, both the biophysical models and plasticity models are relatively small scale, such that modeling populations of neurons quickly becomes non-tractable. Network-level models are mostly limited to correlational relationships, where a pre-stimulation or during stimulation feature predicts stimulation outcome, and where these features are observational and not something we can easily control, such as pre-stimulation functional connectivity[18–24], and during-stimulation responses[25].

Same stimulation, different response

Another complexity, already alluded to above, is that stimulation does not always produce the same effects, likely due to changes in underlying activity and interactions with underlying activity. As mentioned in the previous paragraph, stimulation response can be predicted based on a variety of non-stationary pre-stimulation features. Therefore, providing the same stimulation under different contexts produces different outcomes. In fact, stimulation is used to measure connectivity changes in different behavioral states and before and after stimulation entrainment[26,27]. Stimulation adaptation is also a known phenomenon where stimulation effects diminish with repeated exposures to stimulation[28,29]. Building effective stimulation protocols will likely require modeling this interaction between stimulation and underlying activity and characterizing how this interaction is influenced by different intrinsic states and by interactions with electrical stimulation.

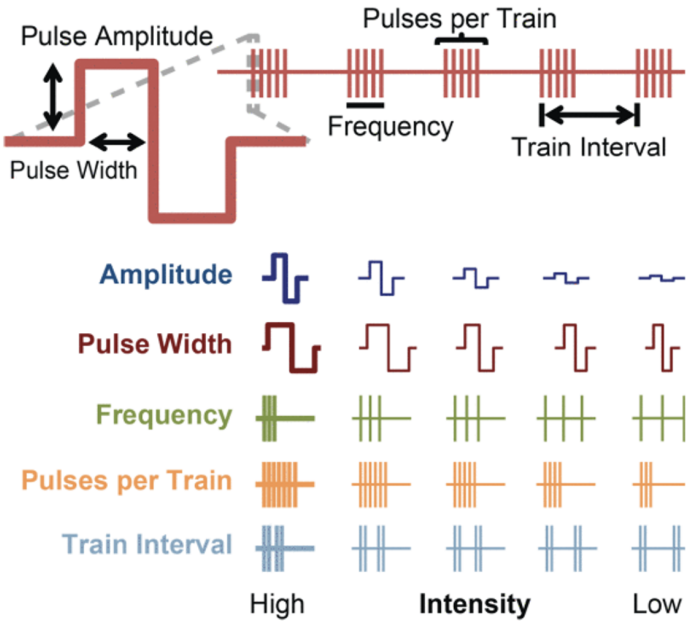


Figure 1.1: (from Bjanes 2019) Summary of five stimulation parameters that were adjusted on a biphasic square wave pulse to explore responses across the stimulation parameter space.

Vast, unexplored stimulation parameter space

A final consideration is that the stimulation parameter space is vast, which provides many avenues of modulation but remains largely unexplored. Stimulation parameters that can be adjusted for individual pulses include pulse shape, pulse amplitude, and pulse width. For sequences of pulses, pulse frequency, number of pulses in a train, and train interval can also be adjusted (Figure 1.1). With these parameters, we can create varying, complex waveforms that change over time. Exploring the stimulation parameter space can potentially uncover more effective stimulation methods, for example in providing a greater range or more naturalistic touch percepts when stimulating the primary somatosensory cortex[30-32]. However, most existing stimulation protocols are either based on protocols that have worked in the past or are borrowed from other modalities, such as applying transcranial magnetic stimulation protocols in intracranial stimulation studies[25,27]. Prior groups have attempted strategic testing of this stimulation parameter space[30,31,33], but the parameter space is far too large to effectively explore and not feasible when working with less obvious behavioral outcomes such as mood or pain. These strategic tests also assume that the same stimulation produces a relatively consistent response, which we know is not necessarily a valid assumption from prior literature. While my thesis work does not attempt to fully explore this parameter space, a

major aspect of my work includes varying stimulation parameters to investigate how changing these parameters changes the stimulation response.

1.3 Emerging research for adaptive closed-loop stimulation

In the last few years, several groups have had some success modeling this relationship between stimulation and neural response in rats[34,35] and in monkeys[36], where contributions include using a linear dynamical system model to explain the temporal relationship and interaction between intrinsic neural activity and stimulation input[34,36], and demonstrating a framework of adaptive closed-loop stimulation[35,36].

Yang et al.[36] were one of the first to create an explicit modeling relationship between electrical stimulation and neural activity by using a Gaussian linear dynamical system (GLDS) model. Using repeated sequences of intracortical electrical stimulation that varied in stimulation amplitude and frequency in non-human primates, they learned a GLDS that modeled neural responses to stimulation using inputs of stimulation amplitude and frequency. However, their model would predict the same neural response given the same stimulation input, where previous time steps of neural response were not considered when making forward predictions. In other words, their model only predicted the "deterministic" response to stimulation, with the understanding that model error included intrinsic dynamics not related to stimulation. Their work presented a proof-of-concept of how their GLDS model could be used to dynamically adjust stimulation to reach a specific neural behavioral target, but such demonstration in vivo has not been achieved in non-human primates or in humans.

Bolus et al.[34] applied optimal control theory to control the firing rate of a single neuron by adaptively adjusting the light intensity of optogenetic stimulation in rats. They similarly modeled their system as a GLDS fitted with experimental data of optical stimulation at various intensities. Then, they defined a constrained linear quadratic regulator cost function to learn the appropriate optical input to produce a set steady-state firing rate. Their combined GLDS model with optimal control was able to produce steady-state firing rates in vivo, demonstrating a closed-loop method to control neural firing rate. This work sought to control individual neural firing rate using a single degree of control (light intensity), and it remains unknown if the same optimal control methods can be applied once scaled up to

population-level dynamics in addition to adding multiple degrees of stimulation input control with electrical stimulation.

Tafazoli et al.[35] introduced a general framework for an adaptive closed-loop stimulation (ACLS) system where stimulation parameters are updated to reach a target neural state based on a model error function. In their work, they defined this error function as the Euclidean distance in a dimensionally reduced space between the current neural state and desired neural state, but they noted that any appropriate cost function can be used in place. A unique contribution from their work was using a "greedy" stochastic learning algorithm to learn the best stimulation parameters to drive the system towards a target state. In contrast to the previous two papers that created GLDS models of stimulation and neural response, this learning algorithm is model-free, where stimulation conditions are iteratively tested and selected to minimize the cost function, which was demonstrated to be robust to neural drift and stimulation adaptation in silico and in vivo in rats. While this model-free learning approach has clear benefits of being robust to neural drift and stimulation adaptation, it is entirely reliant on the defined cost function, requires a lot of training data, and may have limited utility in applications where ACLS requires fast time scales.

1.4 Opportunities for improving adaptive closed-loop stimulation

Brain state estimation for real-time feedback

Building a closed-loop system requires a feedback mechanism to provide information about the current brain state. There have already been significant contributions in identifying indicators of current brain state, or biomarkers, in DBS applications, including in tremor[4,5,37,38], depression[9,39], chronic pain[40], and obsessive-compulsive disorder[10]. These biomarkers can help indicate current symptom state—whether it is a desired or undesired state—and in turn help inform when stimulation therapy is needed. Moving towards patient-specific biomarkers may be needed for applications such as mood disorders and pain, where neural representations of these more complex processes may vary between subjects.

In addition to biomarkers that indicate current symptom state, biomarkers of other brain states, such as arousal or wakefulness, can be useful in informing stimulation therapies as well. Scangos et al.[7] demonstrated that stimulation during different arousal states produced

opposite effects in mood elevation. Stimulation has also been shown to produce different responses based on wakefulness[41,42], although it is unknown whether these different responses also translate to different behavioral outcomes. With examples from prior work, we predict that providing additional biomarker information of "relevant" brain states will improve closed-loop therapies. One challenge is identifying what brain states are relevant and can provide useful information that symptom-based biomarkers may miss. Ultimately, identifying biomarkers of brain state can build upon our current understanding of specific brain states, in addition to informing closed-loop systems for delivering effective stimulation therapy.

Building stimulation response models

With neural biomarkers that estimate current brain state to determine whether therapy is needed, the next step is to program stimulation therapy that drives neural activity from an undesirable state to a desirable state. In the previous section, we reviewed recent efforts in building stimulation response models in rats and non-human primates, where data-driven algorithms were used to adjust stimulation parameters. In humans, stimulation protocols have typically been constrained to threshold-based methods, where upon exceeding a biomarker threshold, stimulation is turned on until the biomarker returns below the threshold[37,43]. While this threshold-based method has been effective in reducing tremor[8,37], there lacks an explicit relationship between stimulation input and biomarker activity. To our knowledge, there has only been one study to date in humans, where Velisar et al. discovered an almost linear relationship between stimulation amplitude and beta power in the subthalamic nucleus, such that modulating stimulation amplitude resulted in predictable changes in beta power[43].

Understanding how changing stimulation parameters may produce varying outcomes is needed for developing effective stimulation strategies, particularly in pain and mood disorders where stimulation-based treatment strategies have been variability effective between patients[44,45]. Being able to predict stimulation response allows researchers to more systematically test stimulation protocols to modulate neural activity towards a more desired state.

Identification of effective stimulation regions

In addition to determining effective stimulation inputs for driving neural activity, there are opportunities to identify effective stimulation regions for neuromodulation. Past methods used for determining stimulation sites include using known anatomical circuitry, lesion

studies[46], functional neuroimaging and tractography[47,48], and opportunistic discoveries from other pathologies[49,50]. Stimulation mapping has also been effective at identifying patient-specific outcomes from stimulation[9], where researchers test individual stimulation sites and patients will report any immediate behavioral changes, which can be a time-consuming task when there are many potential stimulation electrodes.

Biomarker-informed stimulation mapping may be a more efficient approach, where we leverage known biomarkers to help identify stimulation sites that produce changes in biomarker activity. While changing biomarker activity does not guarantee a therapeutic behavioral response, we argue that this expedited process can allow for a more comprehensive initial screening of candidate stimulation regions.

1.5 Contributions to neuromodulation research

The primary contributions of this thesis are as follows:

Review and comparison of stimulation artifact removal methods
Stimulation artifact remains a large barrier to analyzing data during ongoing stimulation. In Chapter 3, we review existing methods used for stimulation artifact removal and identify removal methods that are most appropriate for our stimulation dataset. This work is a crucial prerequisite due to the impact of remnant stimulation artifact and its ability to bias our subsequent research findings.

Building models for brain state estimation
Brain state estimation is of particular importance in closed-loop therapies, where stimulation may only be needed during specific brain states. In Chapters 4 and 5, we demonstrate two examples of brain-state estimation approaches. In Chapter 4, we use an unsupervised semi-hidden Markov model to learn sleep/wake states from several hour long human intracortical data. We show that our model outperforms other unsupervised models and discuss how unsupervised models could be leveraged in situations where data are plenty, but labels are sparse. In Chapter 5, we present efforts in identifying subject-specific biomarkers related to pain and discuss how subject pain reports may influence our ability to learn effective models for pain prediction.

Characterizing local field potential responses to patterned stimulation

Building a stimulation response model first requires collecting stimulation responses and identifying features and regions that can distinguish between different stimulation inputs. In Chapter 6, we introduce an experimental paradigm in humans involving patterned stimulation, where the goal was to identify (1) in what regions did stimulation produce distinct responses and (2) to what extent were these responses distinguishable. We identify that most subjects had few, select channels that were capable of distinguishing between stimulation patterns and that some but not all stimulation patterns are separable. This work builds on our foundational understanding of neuromodulation. We make recommendations for future work to continue towards building a general model for stimulation response.

Mapping regions responsive to neurostimulation

Identification of effective stimulation channels is critical in current and emerging applications of deep brain stimulation. In Chapter 7, we demonstrate that single-pulse stimulation response can be used to predict channel sensitivity to stimulation patterns.
We then propose and demonstrate an experimental paradigm that creates stimulation response mappings, which produces recommendations for stimulation pairs that would be good candidates for modulating a specified channel.

We conclude in Chapter 8 how this work contributes to the broader goal of developing adaptive closed-loop stimulation therapies, and we discuss limitations of current work in addition to recommendations for future work and neuroethical considerations.

2. General methods

2.1 Subjects

Our subjects include patients with intractable epilepsy undergoing week-long seizure monitoring at Harborview Medical Center in Seattle, Washington, USA. As part of routine monitoring procedures, subjects were implanted with multiple intracranial electrodes, typically either stereo-electroencephalography (sEEG) depth electrodes (0.86mm diameter, Ad-Tech Med Instr Corp, USA) or electrocorticography (ECoG) surface electrodes (2.3mm diameter, Ad-Tech Med Instr Corp, USA). Implant regions were determined entirely for clinical monitoring purposes, and as a result, electrode placement varied between subjects. Data collection was approved by the Institutional Review Board of the University of Washington, and all subjects gave written, informed consent.

2.2 Recordings and Data Acquisition

Experimenter collected data

Data were collected using the Tucker Davis Technologies (TDT) PZ5 Neurodigitizer Amplifier and RZ Bioamp Processor devices, enabling collection of up to 128 channels and up to 12 kHz sampling rate. Stimulation experiments were carried out using an additional Subject Interface Module (SIM) device, which connected to the TDT device and enabled simultaneous stimulation and recording.

Clinically recorded data

Clinical data were collected in 24-hour segments throughout the subject's stay, using the Natus Quantum Amplifier device, recording up to 256 channels at 1 kHz sampling rate.

2.3 Experimental protocol

All experiments were performed and collected using the TDT. A brief overview of a typical experimental session can be seen in Figure 2.1. An initial resting state recording was performed to obtain baseline recordings and to help evaluate recording quality and noise levels. After determining stimulation channels, stimulation experiments were performed in

blocks, which included a resting state period, evoked potential measurement, and the stimulation experiment. Short breaks were taken between experimental blocks to minimize lingering effects of previous blocks.

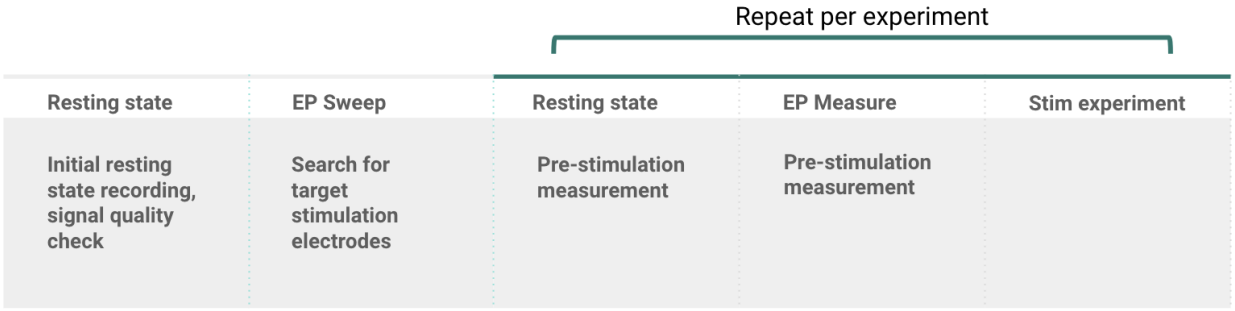


Figure 2.1: Typical experimental protocol during a research session.

Stimulation waveform

Bipolar, biphasic square wave pulses (Figure 1.1) were used as stimulation input with a fixed pulse width of 200 μs/phase. Stimulation frequency and amplitude parameters were varied depending on the experiment, while strictly adhering to current safety limits of not exceeding a charge density of 30μC/cm^2/phase[36].

Stimulation electrode selection

Electrodes near seizure onset regions were not considered for stimulation due to having an increased chance of triggering seizure-related activity. Stimulation electrodes were selected by performing a single-pulse stimulation sweep (amplitude=6 mA, frequency=0.5 Hz) across all available pairs of adjacent electrodes, and we identified stimulation electrodes that produced evoked potential responses across several other channels. The goal was to select a stimulation pair that already demonstrated effective connectivity with other regions, such that use of the stimulation pair in experiments would yield a higher chance of producing measurable responses. If there were multiple candidate electrode pairs, we would preferentially select electrodes that were previously used in earlier subjects.

Evoked potential measurement

Prior to any electrical stimulation experiments, we performed repeated single-pulse stimulation (amplitude=6 mA, frequency=0.5 Hz) at the selected stimulation site to obtain a measurement of effective connectivity[26]. In contrast to resting state connectivity, effective

connectivity is a connectivity measure specifically in the context of stimulation, and it has been used to predict outcomes in subsequent stimulation experiments[26].

Stimulation experiment

Stimulation experiment details are explained in depth in their corresponding chapters.

2.4 Signal processing

Signal cleaning

Channels that had unrecoverable signals due to visually excessive noise or artifacts were removed from analysis. Data preprocessing was performed in the Python environment. Our data cleaning pipeline included a 100 Hz low-pass filter, downsampling to 200 Hz for computational efficiency, a 1 Hz high-pass filter for removal of baseline drift, and a notch filter at 60 Hz to remove line noise. All filters used a zero-phase 2nd order Butterworth filter. In recordings where data during ongoing stimulation will be examined, additional notch filters were applied at stimulation frequencies and their harmonics.

Data analysis

Specific methods and features used will be reviewed in depth in their respective chapters.

2.5 Cortical reconstruction

A pre-operative MRI and post-operative CT brain scan were performed for each subject, and these images were aligned to localize electrodes onto patient brain anatomy. After alignment, we transformed electrode positions from subject space to a standardized brain space created by Montreal Neurological Institute known as "MNI" space[51], which is an anatomical brain map that allows for making comparisons across subjects. The alignment and transformation were performed using the FreeSurfer software package[52].

3. Evaluating stimulation artifact removal methods

3.1 Introduction

One of the biggest challenges in analyzing neural electrophysiology during ongoing stimulation is dealing with stimulation artifact. Stimulation artifacts are often orders of magnitudes larger than underlying neural activity, with a larger presence in recording electrodes closer to the stimulation site. Removing stimulation artifact is required to perform both time-based and frequency-based analysis, as artifact can introduce non-neurological effects that can bias or obscure neural activity. Several methods of removing stimulation artifact have been proposed, but there has not been an adopted standardized method, likely due to the requirement of bespoke solutions for each research group's unique devices, electrodes, experimental design, and research questions.

This chapter includes a literature review on artifact removal methods and a case study where we compare artifact removal methods and identify which method is the most appropriate for our dataset.

3.2 Background

Electrical stimulation is a powerful tool for studying the brain, allowing researchers to activate, suppress, map, and probe neural circuits via activation of neural populations. With numerous clinical applications, including mapping functional regions for more precise tissue removal in patients with medial temporal lobe epilepsy, and deep brain stimulation (DBS) for alleviating tremor symptoms in Parkinson's Disease and in essential tremor, electrical stimulation as a tool has been demonstrated to improve patient outcome, and there is growing interest in improving electrical stimulation methods in addition to exploring its use in different clinical applications. To work towards these research goals, there is an effort to model the effects of stimulation to understand its mechanisms and ways in which we can improve current methods.

To support this type of research, advances in neural recording devices have enabled concurrent sensing and stimulation; however, one of the primary challenges of studying the

effects of stimulation is that it introduces non-biological artifact into recordings. The stimulation artifact is often magnitudes larger than the underlying signal and may distort both or either the temporal or spectral properties of the signal, making removal of this artifact a priority for performing accurate analysis. There are both on-device methods, such as signal blanking, and post-processing approaches to remove or minimize the effect of stimulation artifact on the recorded signal. This review will focus on post-processing methods of stimulation artifact removal that have been used across literature.

3.3 Categories of stimulation artifact removal methods

There are three general approaches to stimulation artifact removal. This chapter seeks to review each method in detail and identify an approach that aligns best with our needs. We provide examples and use cases of stimulation artifact removal methods. A combination of artifact removal methods is also used in some of the referenced papers.

(1) *Signal replacement* - detecting a spike artifact window and replacing the window using interpolation methods[53,54], or with the local field potential (LFP) signal immediately before and after the artifact window to preserve LFP qualities[36,55].
(2) *Template subtraction* - building a template of the stimulation artifact shape and subtracting the template from the signal during stimulation artifact windows[56–62].
(3) *Filtering/outlier methods* - using notch filters at stimulation frequencies[63], Hampel filtering for spectral outliers[64], low pass filtering[65].

Signal replacement methods

Signal replacement methods seek to identify and replace stimulation artifact spikes with another signal, depending on analysis needs. The simplest method is to replace the artifact window with zeros or a mean/median value, which is similar to on-device signal blanking. To add a layer of complexity, linear interpolation[53] or cubic spline interpolation[54] methods have also been used. Another approach is to replace the artifact window with a median-filtered version of the signal, which effectively removes the spike outlier from the signal[66]. In all of these cases, the spike presence is completely removed from the signal; however, these methods should only be applied with time-domain analysis, such as in cortico-cortical evoked

potential (CCEP) analysis since these methods will often introduce distortions in the spectral domain[61].

To preserve the spectral properties of the signal, another signal replacement method involves replacing the artifact window with a signal that has similar spectral properties as the ongoing signal. One method is reversing and concatenating the signal immediately before and after the artifact window[36,55], which preserves the LFP spectral properties while also preventing sudden discontinuities or jumps in the signal, which would also result in spectral artifacts.

Template subtraction methods

Template subtraction methods involve empirically creating a stimulation artifact template and subtracting the template from each stimulation artifact occurrence. This method relies on a key assumption that stimulation pulses, with identical parameters and at the same recording electrode, are very similar or identical in shape. Another requirement for template subtraction methods is that the sampling frequency must be high enough to effectively capture the stimulation pulse shape, where under-sampling will be an issue in both template formation and subtraction steps[59]. The benefit of template subtraction is that when done effectively, it can potentially recover the underlying LFP signal during the stimulation pulse.

Average template subtraction, where an average of stimulation pulses is computed for each channel and subtracted from each stimulation occurrence, has been performed for downstream spike analysis[56,57] and CCEP analysis[58], where downstream methods took place in the time domain. Templates have also been created using function fitting methods[57] and biophysical models[60]. In cases of under-sampling the stimulation pulse shape, a moving average template method has been shown to outperform a stationary average template[59]. Similar template creation methods using a weighted average of the surrounding pulses have also been used[62].

A more flexible template subtraction method involves building a dictionary of templates instead of a single average template, which allows for use of multiple stimulation parameters and performs automatic template matching and subtraction[61].

Filtering or outlier methods

Filtering methods in the frequency domain are also used for removing stimulation artifact, leveraging known information about the stimulation frequency. Notch filters at stimulation frequency + harmonics[63,65] and Hampel filters[64,65] to detect and remove frequency domain spikes related to stimulation artifact have been demonstrated. For stimulation at frequencies higher than the frequency range of interest, for instance, in 130 Hz DBS stimulation with an interest in the 1-100 Hz frequency range, a simple low-pass filter has been used[65]. The limitation of these methods is that signal analysis at or near frequencies contaminated by stimulation artifact may no longer be valid.

In the following sections, we perform a case study that compares the effectiveness of several artifact removal methods on real and simulated human intracranial data with stimulation artifact.

3.4 Methods

Human intracranial dataset

Collection of human intracranial data and general experimental protocols are detailed in Chapter 2: General Methods. We used subject S5 for this case study, where stimulation occurred in electrodes LOF 7-8, which corresponded to the left frontal orbital cortex, with 6 unique stimulation parameters and 10 sequence repetitions (Figure 3.1, Table 3.1). Since different stimulation parameters were used, the stimulation pulse artifact shape varied. Specific to this case study, the only preprocessing step we performed was a high pass filter at 1 Hz used to remove signal drift, which was necessary for several artifact removal methods, particularly median and template subtraction methods, to minimize spike DC offset. No additional filtering or down-sampling was performed prior to artifact removal. A final low pass filter at 200 Hz was used after artifact removal to reduce high frequency noise.

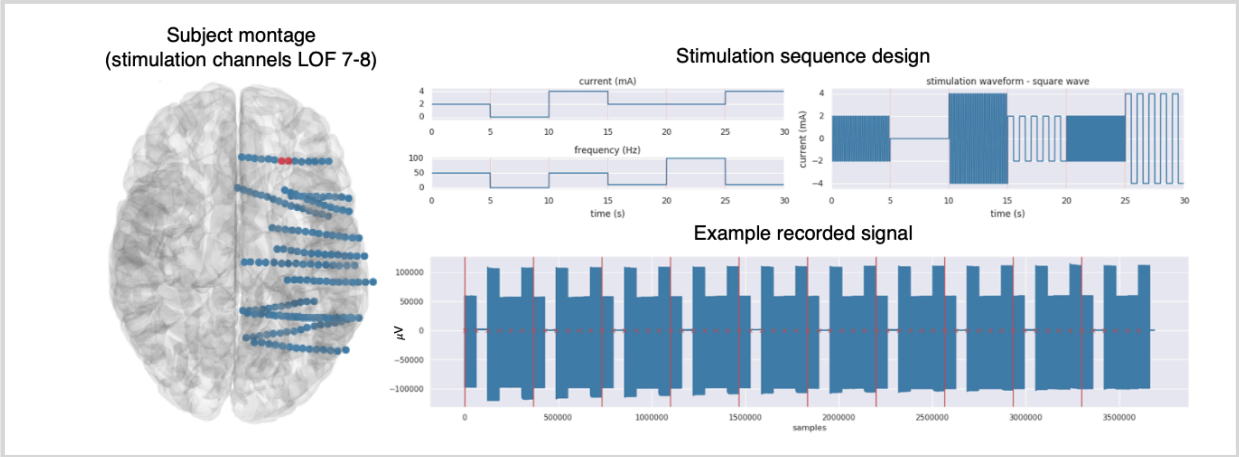


Figure 3.1: (left) Electrode placement in subject S5 with red indicators for stimulation channels (LOF 7-8). (right) stimulation design involved creating a sequence of stimulation inputs that varied in amplitude and frequency, where each input was presented for a fixed duration of 5-seconds. This stimulation sequence was repeated over time, and the stimulation artifact can be seen in the example recorded signal.

Table 3.1: Details of stimulation parameters and simulated sine wave responses from the simulated dataset.

Stimulation condition	Stimulation amplitude (mA)	Stimulation frequency (Hz)	Simulated frequency response (Hz)	Simulated amplitude response (uA)
A	2	50	8	30
B	0	0	0	0
C	4	50	22	80
D	2	10	14	70
E	2	100	42	60
F	4	10	2	100

Generating simulated dataset

One limitation of the human dataset is that the ground truth cannot be completely verified. Therefore, a simulated dataset was generated to mimic neural data from a single channel with stimulation artifact added onto the signal. In addition, we generated artificial stimulation responses, which were increases in predefined frequency bands time-locked to specific stimulation presentations. This simulated dataset was generated using the same data format and experimental design as the human intracranial dataset (Figure 3.2).

The neural data were generated by first using a 60-second period of resting state activity prior to the stimulation experiment. This resting state activity was verified to not include any significant shifts in frequency dynamics. The stimulation artifact was modeled using a reference channel from subject S5, where stimulation pulse timing and its peak value were identified and added to the resting state signal.

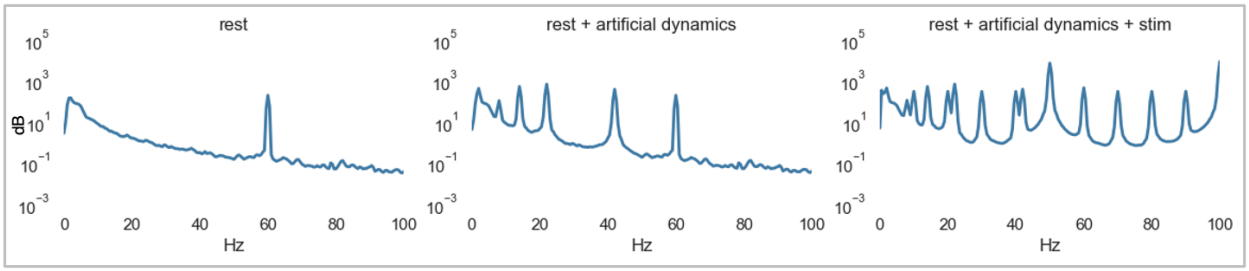


Figure 3.2: Generating simulated dataset using resting state data from subject S5. (left) Resting state power spectral density (PSD) reveals classic 1/f curve with a 60 Hz peak resulting from line noise. (center) Artificial dynamics time-locked to specific stimulation parameters were added to the signal and they appear as distinct peaks in the power spectrum. (right) Stimulation artifact pulses added to the signal create large distortions in the power spectrum.

To mimic a response to stimulation, we added sine waves at pre-defined frequencies with noise during stimulation presentations, such that each sine wave response would only appear during its assigned stimulation pattern (Table 3.1). Frequencies were selected to avoid 10 Hz harmonics, since we won't be able to verify activity at 10 Hz harmonics due to stimulation occurring at that frequency. Amplitudes were selected such that visible peaks were seen in the power spectral density PSD plot following addition of the response signal.

We performed spike detection on a reference channel (LOF3) to identify the timing and amplitude of the stimulation spikes, and we added individual, single sample spikes to the resting state data. Note that the spike shape of actual stimulation pulse comprises more than a single data point, and this method only detects the highest, positive spike component, and does not identify or model the negative peak component of the spike. When we initially added the negative component of the spike, the artifact-related power in the PSD unexpectedly decreased. To best mimic the PSD behavior of the original dataset, we

continued to use only the single positive spike peak. This method was selected to best resemble the variations between stimulation pulses, such that the methods we used to remove stimulation artifact would perform similarly as with the real stimulation data.

Artifact removal methods

Several artifact methods were implemented to compare their effectiveness at removing artifact presence.

Signal replacement with median value: a threshold-based spike detection method was used to identify timing of stimulation pulses (Figure 3.3, top left). We identified that a window of 20 samples centered on the detected spike was sufficient in capturing the entire spike signal. This window was replaced with the median signal value (Figure 3.3, top right).

Average template subtraction: average templates were created by grouping stimulation pulses by the same stimulation input, using the same 20-sample window identified from spike detection (Figure 3.3, bottom). The average template was then subtracted from each stimulation pulse of the appropriate group. From initial visualization, we already observed that there were remaining spike artifacts after template subtraction.

Dictionary-based template subtraction (Caldwell method[61]*)*: in contrast to average template subtraction, this dictionary-based method groups stimulation pulses by shape, creates an average template from identified groups, and then performs template subtraction.

Notch filtering to stimulation frequencies: notch filters were created using 2nd order Butterworth filters with a filter width of 2 Hz centered around stimulation frequencies [10, 50, 100] and their harmonics, which ended up being all multiples of 10 Hz.

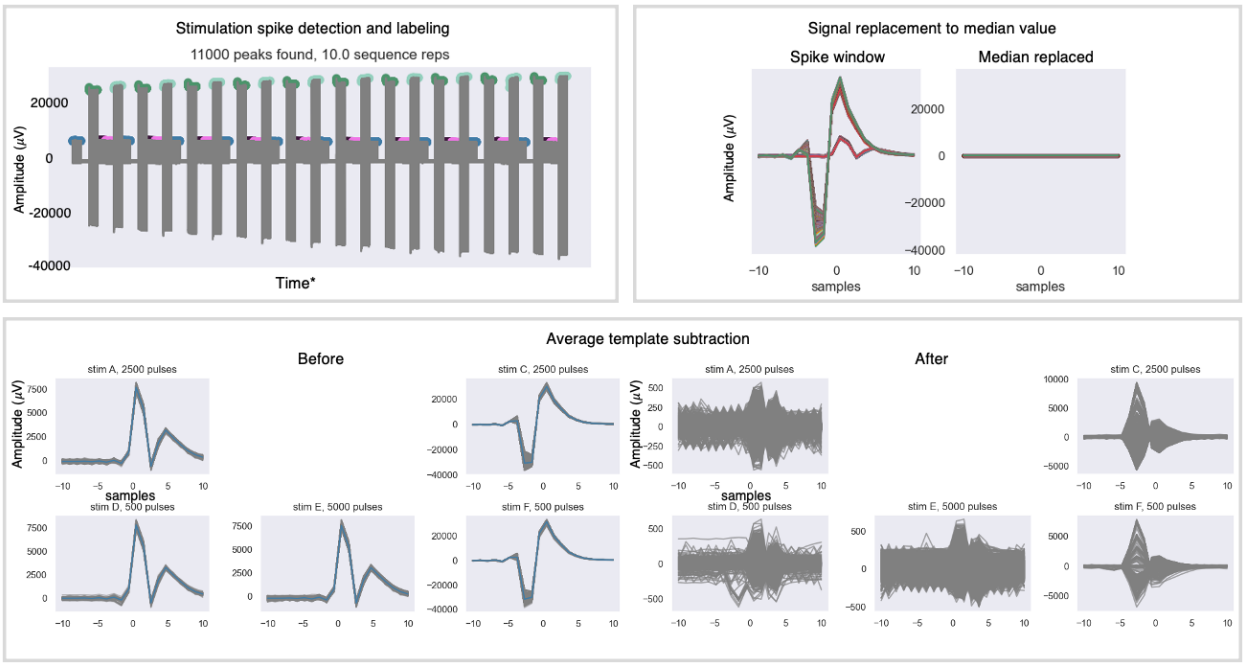


Figure 3.3: Stimulation artifact removal methods. Demonstration of stimulation spike detection (top left), where the total spike count is displayed and the spikes are color coded by stimulation parameter. Signal replacement with median value (top right) was performed using a window size of 20 samples, or approximately 1.6 ms, which was sufficient to remove the entire spike shape. For average template subtraction (bottom), individual templates were created for each stimulation parameter. Spike artifacts are still observed after template subtraction, especially in stimulation conditions C and F.

Evaluating signal recovery

Visualization of the time-signal, PSD, and time-frequency spectrogram provided multiple ways to evaluate the signal quality in comparison to the unfiltered dataset. PSD was calculated using Welch's method[67] with a Hanning window, and spectrograms were calculated using a Tukey window of 500 ms and 1/4th overlap. The resulting spectrogram had a time resolution of approximately 434 ms at frequency intervals of 2 Hz.

We also visualized our datasets after performing dimensionality reduction across the frequency space, as these methods were included in the downstream analysis pipeline. Visualizations were examined for visual discrepancies.

Since we have the ground truth data with the simulated dataset, we additionally reported mean-squared error of the PSD between the filtered data and the ground truth.

3.5 Results

Unfiltered data displayed typical stimulation artifact characteristics

In the unfiltered data, the time signal contained stimulation artifact pulses that were magnitudes greater than the underlying neural signal (Figure 3.4), and these pulses were still present after performing low pass filtering at 200 Hz (Figure 3.5, top row). We noted that the stimulation pulses were not consistent in size, where the first few pulses seemed to have larger amplitudes than the rest of the pulses within a stimulation condition. In addition, the stimulation pulses seemed to be asymmetrical in shape, where in some conditions, the positive peak was at a larger magnitude than the negative peak. This asymmetry is more easily visualized after performing low pass filtering. Finally, we observed that the peak-to-peak amplitude of the stimulation pulses seemed to increase over time, notably increasing in the negative peak component.

The PSD included peaks at 10 Hz and harmonics, with elevated peaks at 50 Hz and 100 Hz that seemed to distort the local 1/f characteristic curve. The spectrogram reflected the temporal structure of the stimulation artifact, where some segments with large 50 Hz artifact seemed to contaminate the entire 40-60 Hz range. Similarly, stimulation at 100 Hz seemed to begin signal contamination at 85 Hz, when compared with the no stimulation condition. Other segments of time displayed dark horizontal bands at 10 Hz increments, which indicated power increases and was aligned with 10 Hz stimulation timing.

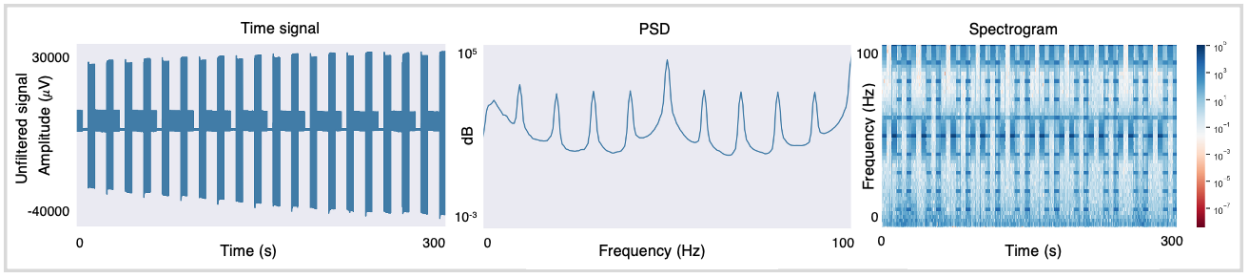


Figure 3.4: Time signal (left), PSD (center), and spectrogram (right) of unfiltered data.

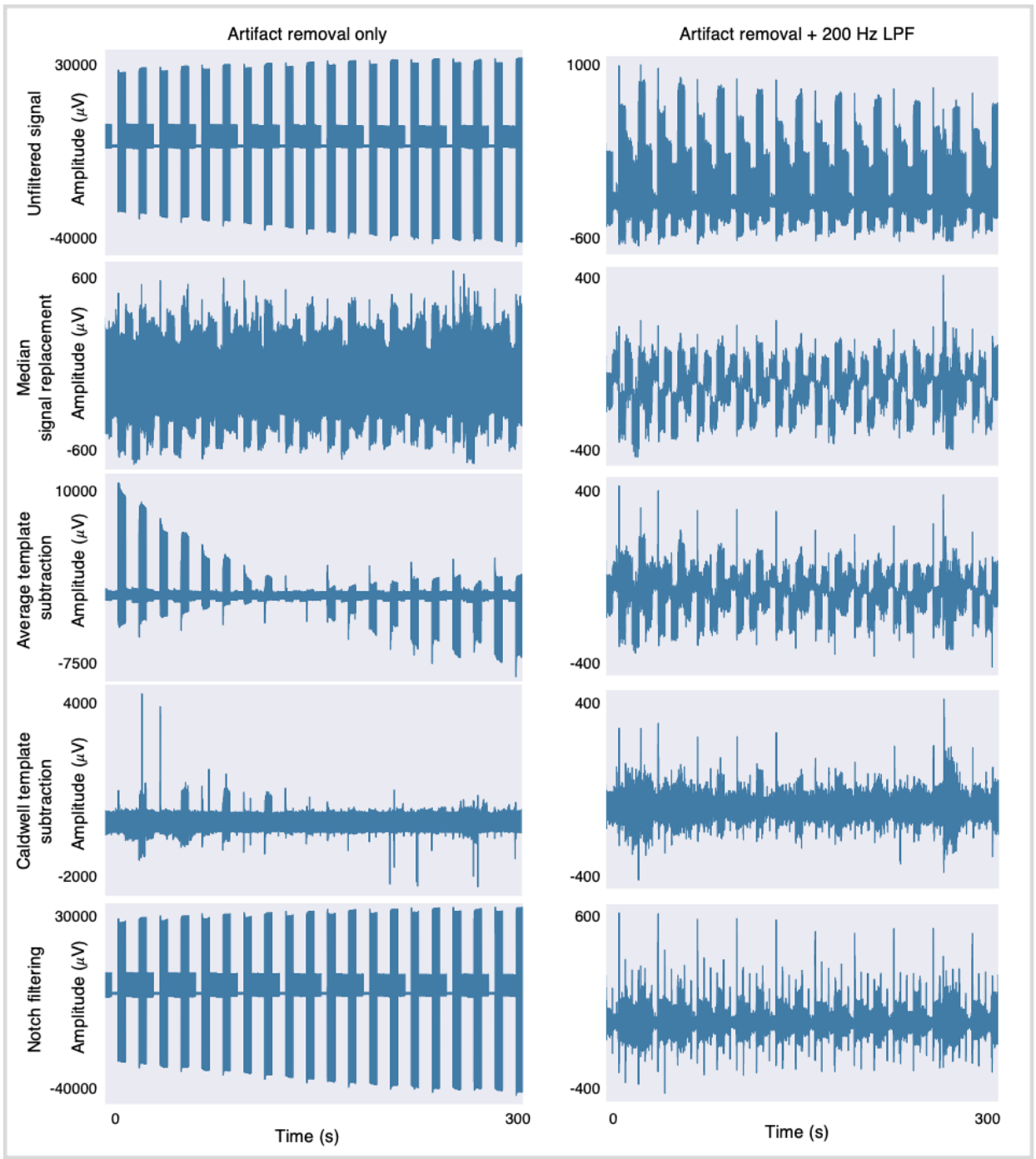


Figure 3.5: Time plot comparisons after artifact removal methods (left) and an additional low pass filter at 200 Hz (right) to remove high frequency noise.

Signal replacement with median value recovered 1/f curve shape

After performing signal replacement with the median value, the underlying time signal seemed to have vertical displacements occurring every 5-seconds, which matched the timing of stimulation presentation. These vertical displacements may be related to the pulse shapes being asymmetrical. In the PSD, peaks at 10 Hz increments were still present but at lowered amplitudes. The PSD shape more closely resembled the characteristic 1/f curve, with much less distortion around 50 Hz and 100 Hz. The spectrogram still had increased power occurring at stimulation frequencies but dramatically improved overall.

Average template subtraction does not handle changing pulse shapes

We observed interesting behavior in the time signal after average template subtraction, where there seemed to be a linear decrease in pulse amplitude centered around zero. This behavior can be explained using our previous observation where the stimulation pulse amplitudes seemed to increase throughout the experiment. Since an average was taken across time for each stimulation condition, the earlier pulses with smaller amplitudes and later pulses with larger amplitudes had poor fits to the average template. The template was most effective towards the middle of the stimulation experiment, and these effects were less pronounced after performing a 200 Hz low pass filter (Figure 3.5, row 3).

Spike artifacts additionally remained, specifically at the beginning of each stimulation presentation. We identified that these remaining artifacts reflected the initial pulses for each stimulation condition that had larger amplitudes than the rest of the stimulation period, where an average template would not be successful.

The 1/f curve in the PSD seemed to improve, although the distortion around the 40-60 Hz range was larger than in median signal replacement. 10 Hz bands in the spectrogram also remained present.

Unsupervised clustering of pulse shapes improved spike suppression

In Figure 3.6, we demonstrated a data-driven method of building average templates by clustering pulse shapes using k-means clustering. We showed that while not perfect, this data-driven clustering method had significantly reduced spike artifacts compared to the average template subtraction method. There were nine distinct spike groups that were

identified, likely reflecting the nine stimulation repetitions performed. Spike shapes after performing template subtraction using the k-means identified clusters produced a resulting spike dataset with much smaller peak magnitudes than in average template subtraction, where spikes may not have been suppressed fully but were significantly improved.

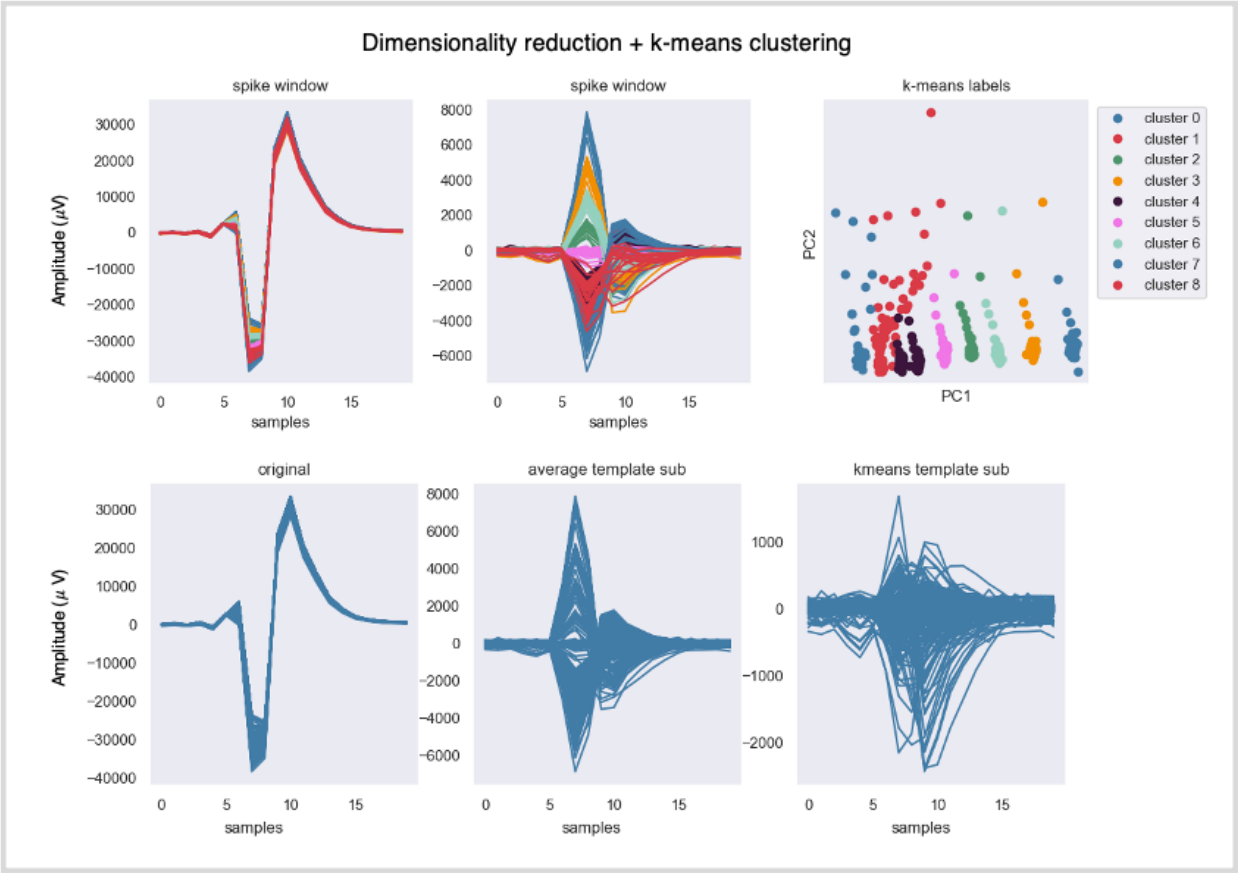


Figure 3.6: Demonstration of k-means clustering on spike shapes to identify clusters for template subtraction. Original spike data (left) and average template subtraction spike data (center) were color-coded based on identified grouping from k-means clustering. Clustering was performed on original spike data only, and plotting cluster results on spikes after average template subtraction helped to better visualize the distinction between spike shapes. Identification of groupings (top right) revealed nine distinct clusters, reflecting the nine stimulation repetitions performed. Spikes after k-means template subtraction had improved spike suppression compared to average template subtraction (bottom right).

Dictionary-learning template subtraction largely removed stimulation spikes
Data-driven learning methods can be leveraged to improve upon average template subtraction, where instead of assuming that all stimulation pulses are the same shape, pulse shapes are grouped together based on similarity, and these groups form data-driven templates.

In the time signal, large amplitude spikes still remained in the recording, but after performing low pass filtering, these large amplitude spikes were removed. In addition, the resulting time signal did not have the vertical displacement behavior seen in the median replacement and average template subtraction methods.

When observing the PSD and spectrogram, the increases in power at 10 Hz increments were still present but at a noticeably decreased power, and there also seemed to be less distortion surrounding the 50 Hz stimulation artifact.

Notch filtering aggressively removed power in stimulation-related frequencies
Due to our signal of interest residing in the 0-100 Hz range, we performed notch filtering up to 100 Hz, which meant that power in all stimulation-related frequencies from 100 Hz to our sampling rate of 12 kHz was still present and reflected in the time signal. A 200 Hz low pass filter was required to properly visualize the time signal after notch filtering (Figure 3.5). The time signal was largely recovered, although some stimulation conditions seemed to have different amplitudes of activity compared to others. Spikes at stimulation transitions were observed, and we believe that these are produced due to a change in stimulation charge delivered. These spikes are similarly seen in the other artifact removal conditions but are less obvious. In subsequent analysis, we removed a 0.5 - 1 second window surrounding stimulation transitions to remove this behavior.

The PSD plot displayed sharp negative peaks at 10 Hz increments, reflecting the notch filters performed on the dataset. The 1/f curve did not seem to be restored, with a large increase in power surrounding 50 Hz and 100 Hz. The spectrogram displayed very little presence of stimulation artifact, where faint white lines ran across the spectrogram at 10 Hz increments reflecting the notch-filtered stimulation frequencies.

Dimensionality reduction revealed effects of still-present stimulation artifact

Dimensionality reduction using PCA was the subsequent procedure in our typical analysis pipeline, therefore, we compared the results from dimensionality reduction across the artifact removal methods used (Figure 3.7). In the unfiltered data, we clearly observed distinct groupings of stimulation response, where these groupings become less distinguishable depending on the artifact removal method used, indicating that the first two PC components represented the variance explained by stimulation artifact. Of the artifact removal methods, notch filtering and the dictionary template method seemed to reduce stimulation group distinguishability the most, which could be attributed to their effectiveness at removing stimulation artifacts. Notch filtering visually appeared the most different, and we attributed this difference to notch filtering being the most aggressive filtering method and removing a significant portion of frequency data from the dataset.

Figure 3.7: Comparison of stimulation responses distributions after each artifact removal method in independent PC spaces, including the original, unfiltered data. Unfiltered data displayed distinct clusters of responses if no stimulation artifact removal was performed. Median replacement and template subtraction had visibly separable clusters of responses, which we posit are a result of the remnant presence of stimulation artifact. Caldwell method

and notch filtering, which had the least presence of stimulation artifact, seemed to have the most overlap in stimulation response across groups.

Artifact removal methods in simulated dataset perform similarly to real dataset

Due to there being no ground truth for this dataset, it is difficult to know at what point the stimulation artifact has been removed where it no longer influences our analysis to a significant extent. We created a simulated dataset using resting state data and artificial stimulation response to create a ground truth, on which we added stimulation artifact. We then tested each stimulation artifact removal method using this new dataset and compared how closely the PSD resembled the ground truth (Figure 3.8, 3.9).

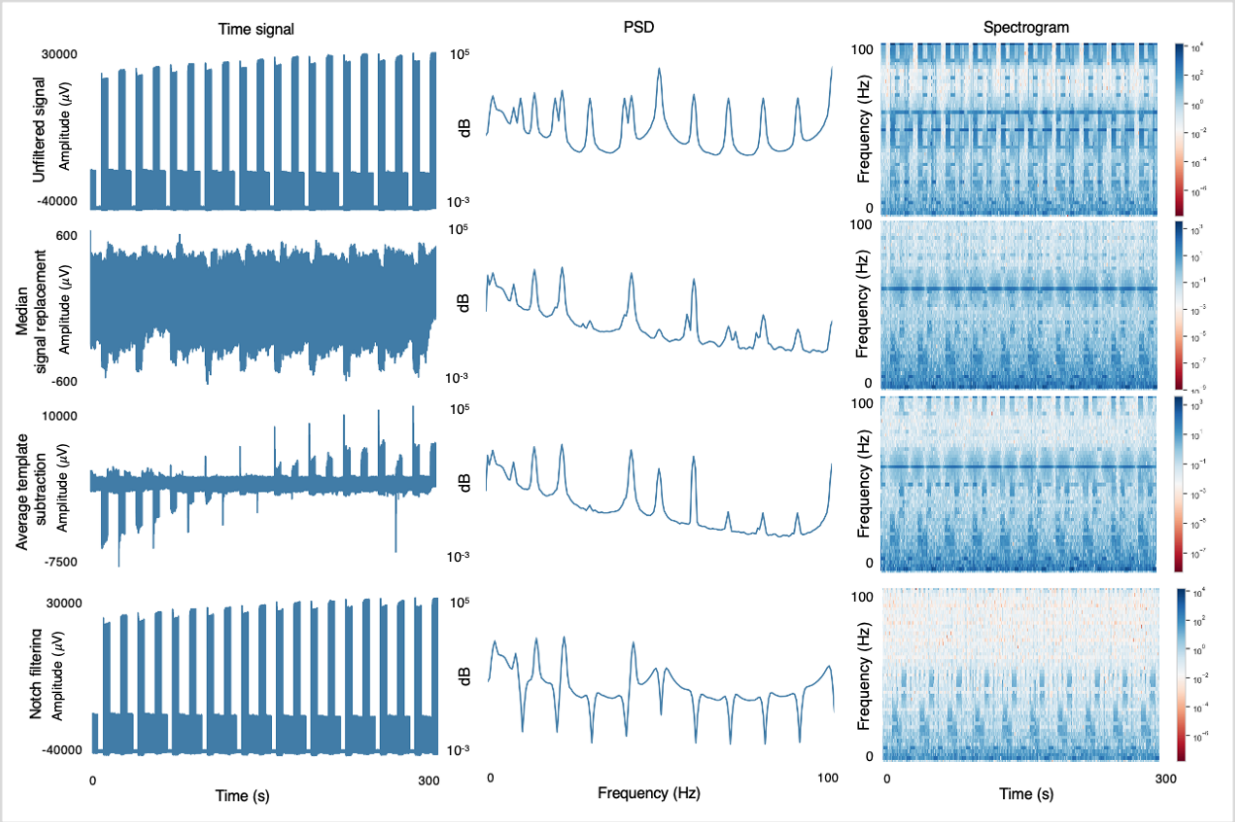


Figure 3.8: Artifact removal methods compared across simulated dataset.

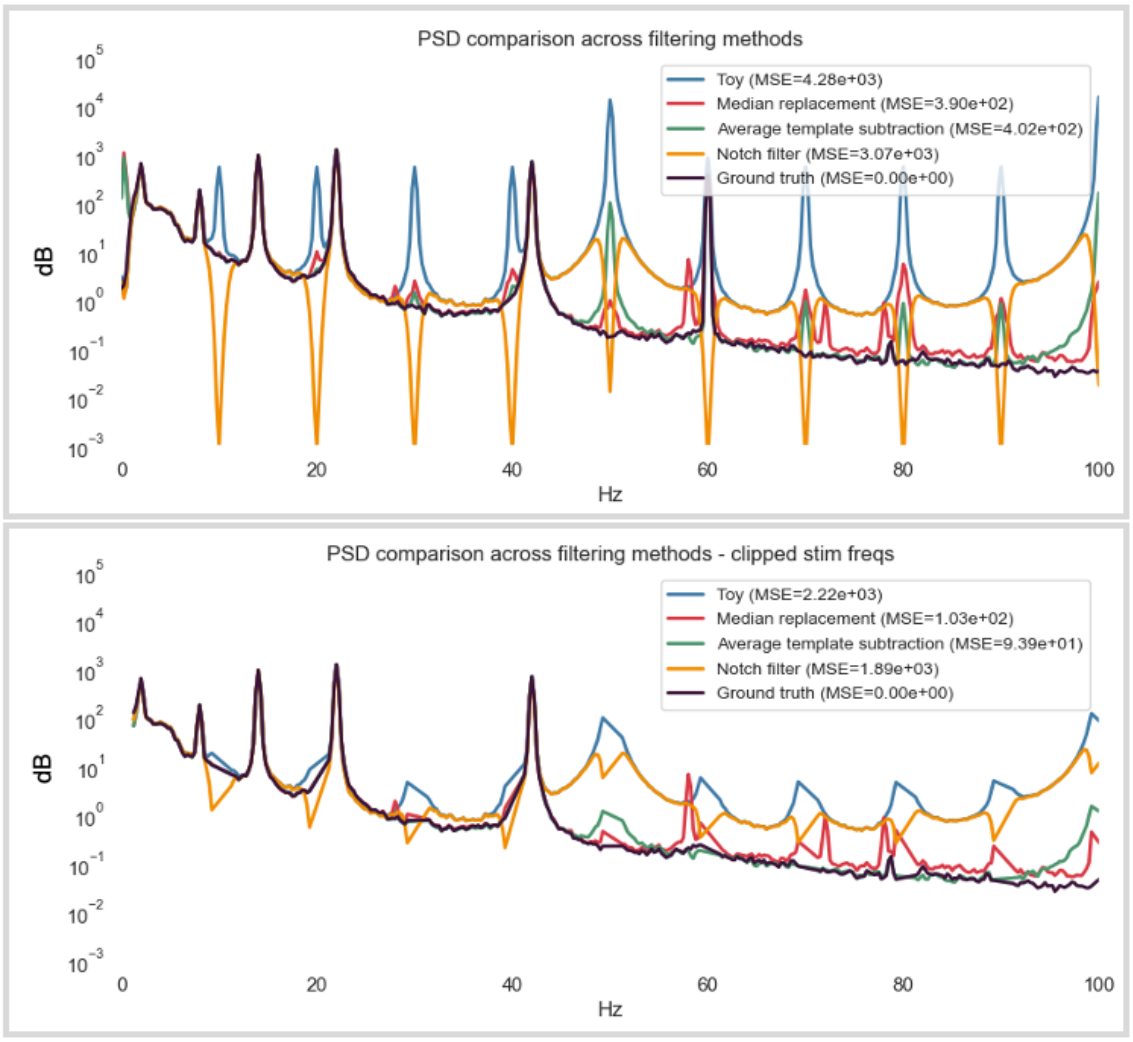


Figure 3.9: Artifact removal PSD comparisons using simulated dataset. Reported mean-squared error (MSE) indicated that median replacement most closely fit the ground truth. In the bottom plot, power information was removed at stimulation frequencies as data at those frequencies would not be used for subsequent analysis. In this scenario, average template subtraction had the lowest MSE compared to the other artifact removal methods.

Median replacement outperformed average template subtraction and notch filtering at restoring the PSD, due to average template subtraction having similar challenges with variable stimulation artifact amplitudes, and notch filtering's large distortion of the PSD (Figure 3.9, top). Since stimulation frequencies would not be included in analysis, we additionally clipped a 2 Hz window surrounding stimulation frequencies and removed this data from analysis (Figure 3.9, bottom). In this context, we observed that average template

subtraction outperformed median replacement and notch filtering with having the lowest MSE compared to ground truth.

3.6 Discussion

Removal of stimulation artifact is crucial in recovering the underlying activity and is an important step in enabling closed-loop response neurostimulation therapies.
Similar to signal processing, one important consideration is that the more aggressive the stimulation removal method is, the more at risk we are at distorting the dataset, which may have potential downstream consequences if not used carefully. A critical decision is determining at what point the filtering and artifact removal is sufficient for our needs. In our case, we saw that stimulation artifact did create an artificial separability between our stimulation conditions, due to our classifiers learning the behavior of the artifact itself. Our goal was then to minimize stimulation artifact, such that the magnitude of neural behavior response to stimulation was greater than the magnitude of stimulation artifact itself. Since we have yet to identify a ground truth in neural behavioral response, it remains a challenge of whether the differences in responses we observe are due to stimulation artifact or due to legitimate response.

To address these challenges, we created a simulated dataset with known behavioral responses and added stimulation artifact from recorded data to simulate our dataset. Using a simulated dataset allowed us to measure differences between ground truth and signals after stimulation artifact removal techniques were used. In addition, we recommend that avoiding data at or near stimulation frequencies would assist in determining the best artifact removal method, as stimulation-related frequencies should be removed from downstream analysis entirely.

The most effective artifact removal method may depend on the downstream features intended for use. Cortico-cortical evoked potentials (CCEPs) are commonly characterized by time-domain features, such as peak-to-peak amplitude and latencies for specific CCEP components. Artifact removal methods for CCEP work may focus more on preserving time-domain features and less on preserving or minimizing distortion in the spectral domain. For example, interpolation methods may introduce spectral domain artifacts and will likely not match the spectral qualities of the surrounding neural LFP signal, and subsequent spectral

analysis may result in bleeding of this artifact. In contrast, power-in-band analysis involves spectral domain features, such that the artifact removal method may focus on minimizing distortion in the spectral domain.

For our dataset, we identified that the most aggressive method, notch filtering, would likely be the best candidate for our subsequent analysis due to our over-cautiousness of ensuring minimal influence of stimulation artifact. However, due to the continued distortion present at 50 Hz and above, we also constrained our frequency analysis to only include frequencies up to 40 Hz.

4. Unsupervised detection of sleep/wake states

Stimulation therapies are known to produce variable results during different brain states, and particularly between sleep and wake states[41,42]. Sleep and wake are of particular interest to closed-loop therapies, where therapies may not be needed during sleep. This chapter demonstrates an example of leveraging machine learning models to predict brain states.

The chapter includes the published work as cited below and was supported by the National Science Foundation (1630178 and EEC-1028725) and the University of Washington Big Data for Genomics and Neuroscience Training Grant (1T32CA206089-01A1).

[published IEEE EMBC 2020] Sun S*, Jiang LP*, Peterson SM, Herron JA, Weaver K, Ko A, Ojemann J, Rao RPN; Sleep/wake classification using a sequential state space model.

4.1 Introduction

Sleep remains an elusive topic in neuroscience, having intricate relationships with learning and memory consolidation[68]. Disturbances of normal sleep patterns frequently emerge and interact with other medical conditions, such as epilepsy[69,70] and many mental disorders[71], and the reasons and effects are poorly understood. Studying the dynamic interactions between sleep and neurological impairments can aid towards building a comprehensive understanding of the disorders and improve patient outcomes.

To study sleep, researchers commonly perform overnight electroencephalography (EEG), and trained technicians label the different sleep stages: N1, N2, N3, and REM[72]. Manual sleep staging involves visually examining 30-second segments of neural data and labeling each segment based on established sleep scoring rules. These rules include decision based logic and identifying specific neural features. A key challenge for sleep staging is that some rules are subject to interpretation, leading to labeling differences between technicians[73]. These rules were also developed based on healthy adult subjects, allowing little extrapolation to other age groups or to those with atypical sleep patterns.

With rapid developments in machine learning, automated sleep staging can resolve these issues by mitigating human bias and providing flexibility to variable sleep patterns. Current efforts in automation primarily use EEG in combination with supervised or unsupervised machine learning models[74-76]. However, little work has been done with electrocorticography (ECoG), with increased spatial resolution and signal-to-noise ratio to capture and reveal subtle brain dynamics related to sleep. Kremen et al.[76] demonstrated using a single intracranial electrode to classify Wake, N2, and N3 stages with 94% accuracy. Their unsupervised model used decision trees with numerical thresholds that required tuning from pre-existing sleep stage-labeled data, which may not always be available. Questions remain regarding whether a threshold-based approach is generalizable to variable sleep patterns. We believe that leveraging sequence-based methods can overcome these limitations and contribute to automation tools for sleep staging.

One nuance is that our ECoG recordings are of patients with intractable epilepsy, a neurological disease known to have complex reciprocal interactions with sleep and often presents with disrupted sleep patterns[69]. Successful development of a sleep stage classifier in these patients must be robust towards variable sleep dynamics between subjects. An unsupervised approach is ideal due to self-learning of features with high distinguishing properties, providing flexibility to differences in sleep patterns or electrode placement. Unsupervised methods also alleviate the impracticality of labeling sleep stages in every patient for model training.

In this paper, we built an unsupervised hidden semi-Markov model (HSMM) to label sleep and wake states in epilepsy patients, using spectral power features from a single ECoG electrode. We show that HSMMs produce higher labeling accuracy and fewer extraneous transitions than k-means clustering and hidden Markov models. Our results provide initial confirmation of HSMM as an effective tool for automated sleep staging, and motivate future applications of HSMM to neural behavioral state classification.

4.2 Methods

Data Collection

Our data consist of continuous neural recordings from patients at Harborview Medical Center (Seattle, Washington) undergoing clinical monitoring for intractable epilepsy. As part of their monitoring procedure, an 8x8 electrocorticography (ECoG) grid (2.3 mm exposed diameter, Ad-tech Medical, Racine, WI, USA) was implanted on the patient's cortical surface, providing 24-hour continuous ECoG data. Data collection was approved by the Institutional Review Board of the University of Washington, and all subjects gave written, informed consent. We sampled approximately 4-14 hour continuous segments of ECoG data from two patients over three days and one day from two additional patients, ensuring that there was no seizure activity during the recording. Sleep and wake labels were coarsely annotated based on patient room video monitoring and synchronized to neural data using timestamp information.

Our signal processing pipeline is summarized in Figure 4.1. We minimally processed the ECoG signal to remove noise and baseline drift, including band-pass filtering 1-200Hz, notch filtering at noise harmonics (60Hz, 120Hz, 180Hz), and down-sampling to 500Hz. Electrodes with excessive noise or signal artifacts were removed from analysis. We selected the first grid channel, typically near the superior frontal gyrus, across patients for clustering to minimize computational cost. We split the continuous data into 14-second epochs with their corresponding sleep/wake labels. Total sleep and wake times are summarized in Table 4.1.

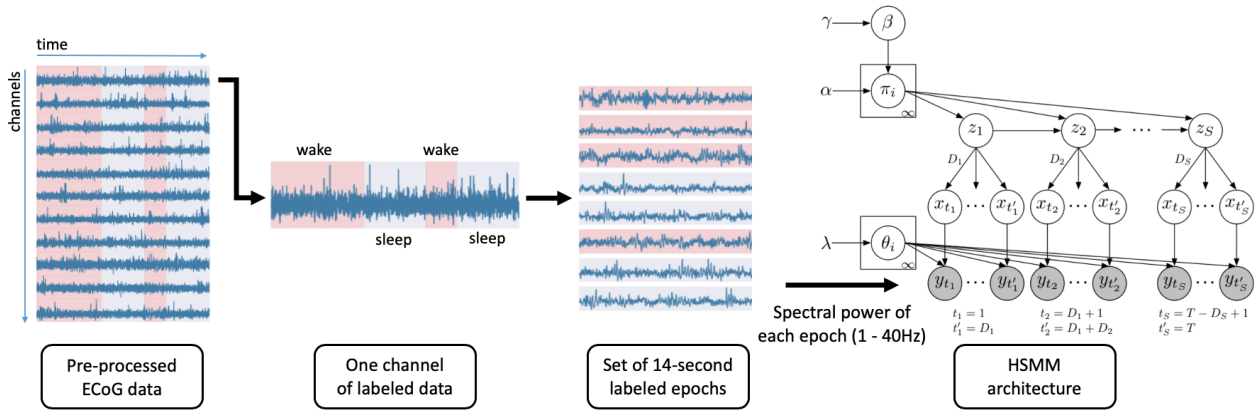


Figure 4.1: Data processing pipeline: ECoG data were pre-processed (bandpassed 1-200Hz, notch filtered at noise harmonics, downsampled to 500 Hz), and a single channel was used for sleep/wake clustering. Data from this channel were epoched to 14-second segments with corresponding sleep/wake labels. Spectral power of each epoch (1-40Hz) was used as input to

the HSMM shown on the right. HSMM Architecture (adopted from Johnson et al.)[77]: β is drawn from GEM(γ), α and β define the Dirichlet process (DP) prior, π represents the transition probability drawn from DP(α,β), λ represents a uniform Gaussian prior for observations, and D represents a Poisson prior for duration states. x and y represent the states and observations. See text for details.

Table 4.1: Subject sleep/wake total times, with total number of epochs (t=14s) also reported.

	Subj 1	Subj 2	Subj 3	Subj 4
total wake time (hr)	25.2	9.5	4.6	5.5
total sleep time (hr)	17.7	11.1	3.5	2.7
total epochs (n)	11053	5295	2074	2117

Feature Selection

Neural analysis commonly involves decomposing the time-series signal into its frequency components due to the association between specific frequency bands and neural activity[78]. Sleep and its progression through different sleep stages are often characterized by shifting towards more low frequency rhythms. For the purpose of sleep/wake clustering, we used frequencies 1–40 Hz to capture a broadband range of frequencies, which we believe is sufficient for distinguishing between wake and sleep. We calculated the spectral power from each integer frequency within that range using Welch's method[67] across epochs. These vectors of length 40 were the inputs to our unsupervised models.

Model Architecture

Sleep has sequential staging with distinct spectral properties, so computational models that leverage sequential properties may better model sleep patterns. We propose an unsupervised implementation of hidden semi-Markov models[79] (HSMMs) as seen in Figure 4. Compared to hidden Markov models (HMMs), HSMMs allow hidden states to have various lengths instead of single-unit transitions in HMMs. Given that individual sleep stages can range from minutes to almost an hour, and sleep and wake periods commonly last for extended periods of time, we believe HSMMs are most suitable for our sleep/wake classification task. We additionally implemented K-means clustering and HMMs to compare across other unsupervised methods.

To determine the prediction label of each cluster, we followed the purity measure[80] of clustering performance – each cluster was assigned the label having the most counts within

that cluster. We evaluated and reported accuracy across our unsupervised methods and across subjects.

1) *K-means Clustering*: We first ignored the sequential nature of the data and used the performance of k-means clustering as a baseline. k-means finds the best k points that minimize the overall Euclidean distance between our data and the corresponding centroids, and in our case, we set there to be two centroids to represent the clustering of sleep and wake epochs.

2) *Hidden Markov Model*: Our HMM was initialized with two states (sleep, wake) with transition probabilities of 0.9 (self) and 0.1 (change). Emission probabilities were set as a multivariate Gaussian, and transition and emission parameters were updated using expectation-maximization (EM). We used Viterbi inference[81] to produce the best state sequence for the expectation step, and maximum likelihood estimation (MLE) to update parameter values for the maximization step, iterating until convergence. We ran Viterbi inference one final time to produce the predicted sleep/wake sequence.

3) *Hidden Semi-Markov Model*: We followed a Bayesian non-parametric approach[77] and used a Dirichlet process (DP) prior for sampling the transition probability between states and a Poisson prior for the state duration length. Figure 4.1 shows a graphical representation of the model. To keep the computational time tractable, we limited the maximum number of states the model can use to four while allowing the model to select the ideal number of states. We set the conjugate prior to Gamma(1000,5) for inferring the Poisson state duration. We used $\gamma = 0.6$, $\alpha = 0.6$ for Dirichlet process priors. The parameters were estimated through posterior samples from Gibbs sampling[82] for 150 iterations. While we did not define a convergence criterion, the model converged after an average of 110 re-samplings.

4.3 Results

Figure 4.2 summarizes sleep/wake clustering performance across our three unsupervised models on one day of data from Subjects 1 and 2. As expected, k-means clustering resulted in the lowest classification accuracy of 59.6% (Subject 1) and 57.4% (Subject 2). HMM performed better with an accuracy of 87.3% (Subject 1) and 75.1% (Subject 2). The HSMM performed best

out of the three, with an accuracy of 90.4% (Subject 1) and 75.5% (Subject 2). While the HSMM accuracy was only slightly higher than the HMM accuracy for Subject 2, the HSMM visually provided more accurate transitions between sleep and wake states. We attributed these infrequent transitions to the additional duration component in the HSMM, which seemed to create more strict requirements for transitions to occur. However, these stricter guidelines may explain why the HSMM did not capture some short-duration wake states.

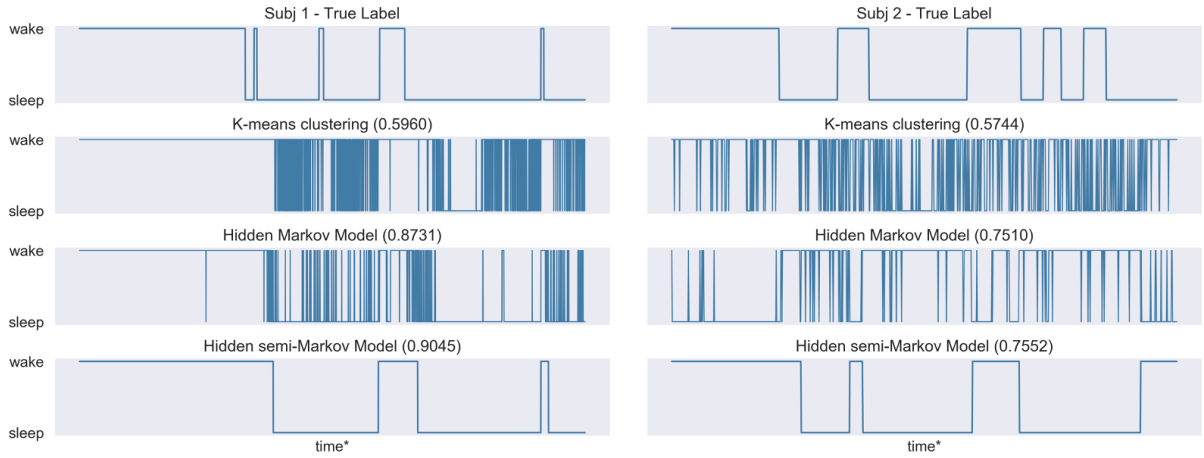


Figure 4.2: Subject 1, 2 sleep/wake true labels and clustering comparison across three methods over one 24-hour day, starting at approximately 8am. Model accuracy for Subject 1 and 2 are shown in parentheses next to the model used. HSMM produced the highest accuracy across models and minimized unnecessary transitions.

When comparing model performance across all days for all subjects (Table 4.2), HSMM produced the best accuracy except in Subject 4. All models tended to perform worse for Subjects 3 and 4. Visualizing the HSMM performance across subjects (Figure 4.3), we noticed that while the HSMM captured the general pattern of Subjects 1 and 2, albeit missing some short duration transitions, it poorly matched the pattern for Subjects 3 and 4. One explanation for variable performance across subjects is that we froze model parameters, and these parameters may perform better in subjects with longer state durations than in others. These results suggest that generating subject-specific parameters is necessary to better handle variable sleep patterns.

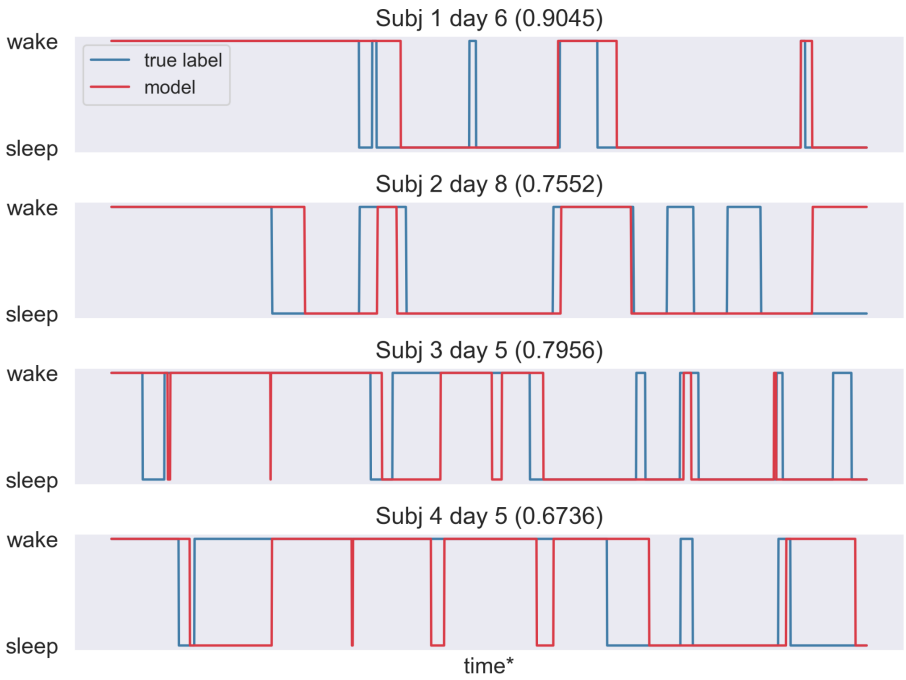


Figure 4.3: Intersubject comparison of HSMM performance in classifying sleep/wake states over one 24-hour day, starting at various times. Performance varies across subjects, capturing major sleep/wake transitions but failing to capture some short-duration states.

Table 4.2: Accuracy of unsupervised models across individual subjects and overall (weighted on subject data length)

	Subj 1	Subj 2	Subj 3	Subj 4	Overall
k-means	0.7331	0.7014	0.6451	0.6476	0.7217
HMM	0.8523	0.8042	0.7623	0.6840	0.8146
HSMM	0.8939	0.8410	0.7956	0.6736	0.8517

To determine if electrode selection was critical in HSMM performance, we ran the model for one day of Subject 1 using each electrode individually and reported model accuracy (Figure 4.4). Accuracy ranged from 60.2% to 97.5%, with more than half the electrodes having greater than 90% accuracy. The lesser performing electrodes were located along or just superior to the Sylvian fissure, which suggests a regional dependence of whether the electrode is a good candidate for model input.

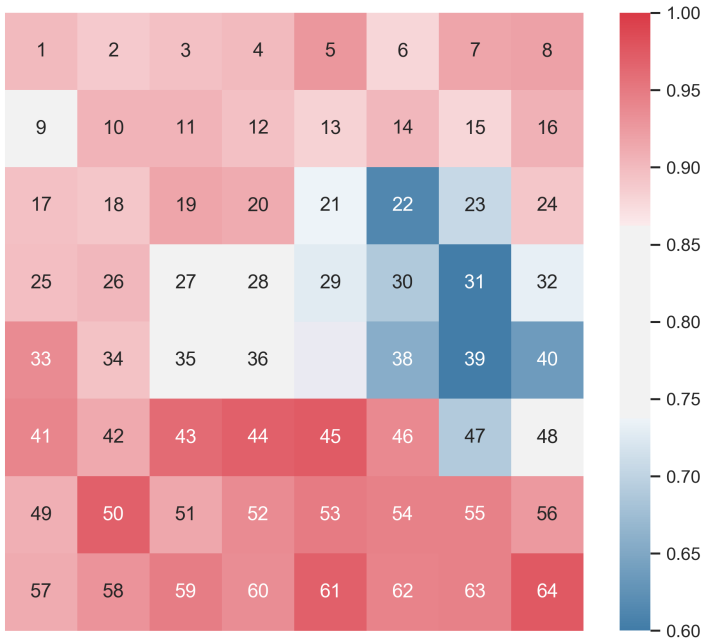


Figure 4.4: Comparison of HSMM performance when using different ECoG electrodes (1-64) in Subject 1. Color bar denotes prediction accuracy. Most electrodes produced an accuracy greater than 90%, while a smaller set in the center-right produced lower accuracies. Electrode 37 was omitted due to excessive signal artifacts.

4.4 Discussion

We demonstrated a Bayesian non-parametric HSMM approach for unsupervised labeling of sleep and wake states while incorporating smooth state transitions. While HSMM performance varied among subjects, HSMM consistently performed best across the unsupervised methods. We also showed that in one subject, HSMM accuracy remained consistent when using different electrodes, noting that electrode location had considerable importance to model performance and should be further explored. These results serve as a benchmark of using a Bayesian non-parametric approach for unsupervised classification of behavioral states in ECoG.

One limitation of using advanced machine learning models is that the abstraction tends to shift away from biological interpretability. The benefit of HSMM is that it explicitly models a duration distribution, which enables us to provide prior knowledge for the duration of sleep and wake segments. Such an approach is highly interpretable in terms of biological significance but may require modification to account for variability in sleep patterns for successful state classification across a variety of people.

With the goal of automating sleep staging in ECoG, future work will employ this HSMM on data with sleep stage labels. The next step would be to stratify the sleep states for further clustering into sleep stages. Since this dataset is uniquely from epilepsy patients, additional seizure state identification is critical to ensure seizure events are classified separately. Due to the high comorbidity between epilepsy and impaired sleep, labeling both sleep stages and seizure events could provide new insights into the interplay between sleep and epilepsy and translate clinically as a tool for seizure labeling, prediction, and localization. Improving tools in automated behavioral state classification will allow broader application to other neurological states and impairments and will open possibilities in characterizing a range of behavioral states and their interactions.

In the context of building adaptive stimulation therapies, we envision that this unsupervised sleep/wake model could provide useful brain state estimates that inform stimulation programming. A simple use case could be using this model to ensure that stimulation therapy remains off during sleep in tremor patients, when active stimulation therapy may not be needed. Another use case could be for epilepsy patients, where in some cases, seizures are more likely to occur during different stages of sleep[69]. Since this model predicts a general behavioral state rather than a symptom-related state, its current model accuracy range of 67% - 89% between subjects may be sufficient in providing contextual information to the model. Due to the model's current performance, additional safeguards should be put in place to prevent potential repercussions caused by inaccurate predictions, such as turning off therapeutic stimulation due to false detection of sleep. Similar unsupervised models have achieved prediction in other potentially relevant behavioral states, such as movement and speech in naturalistic human iEEG[83], and next steps include identifying how these behavioral states may influence stimulation outcomes in addition to how we may leverage these contextual biomarkers to improve clinical outcomes.

5. Biomarker identification of acute pain in humans

In the previous chapter, we discussed building models to predict current brain state due to its pertinence to stimulation therapies. In this chapter, we focus instead on predicting symptom state, where a model can serve as a real-time indicator of a specific symptom. We specifically seek to identify patient-specific neural biomarkers of pain, as pain has been characterized as having multiregional interactions[84], and there is growing interest in neuromodulation applications for chronic pain[85]. Identifying neural biomarkers of symptom state can provide valuable insight into underlying mechanisms in addition to informing adaptive stimulation therapies of when therapy is needed.

In this chapter, we summarize work completed in collaboration with Timmy Vu Pham. The work presented here is a subset of the total work completed by Timmy as part of his master's thesis[86]. My contributions to this work included project conception, planning, and overseeing progression. Timmy completed the analysis and figure making. All text in this chapter was written by me.

5.1 Introduction

Pain is an important, complex process that alerts individuals to physical or perceived harm, where the IASP definition of pain[87] is "an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage." Prolonged pain that transitions into chronic pain can be debilitating to one's quality of life, and ongoing research seeks to identify effective treatments for chronic pain.

Functional imaging has revealed a pain matrix[84], or a collection of brain regions involved in pain processing including sensory, motor, and cognitive processing areas. It is known that pain is not only a physical, sensory experience, but also an emotional experience, which translates to several brain regions acting in parallel to produce the conscious experience of pain.

Current treatments for pain are not sensitive to the complexity and multidimensionality aspects of pain. While effective, pharmacological treatments such as opioids have high rates of misuse and addiction[88]. There has been growing interest in applying neurostimulation

interventions such as deep brain stimulation (DBS) for treatment of chronic pain, with some initial success in well-selected patients where pain mechanisms were known[44]. Closed-loop DBS for pain has been proposed as a feedback-based treatment option, where estimating current pain state using biomarkers can inform treatment strategies[89]. Shirvalkar et al examined long term pain biomarkers in chronic pain patients and identified promising biomarkers in the orbitofrontal cortex[40]. Our research extends this initial work in humans, where we leverage the widespread electrode coverage in epilepsy patients undergoing seizure monitoring to conduct a more comprehensive search for pain-related biomarkers.

We aimed to identify subject-specific biomarkers of pain and to evaluate biomarker performance in the context of acute pain. Our dataset included week-long intracranial encephalography (iEEG) recordings from five epilepsy patients, in addition to subject-reported pain levels on a 0-10 visual analog scale (VAS). Neural features, including power-in-band and connectivity features, were used as input for model trainings, where an array of models was evaluated for pain classification performance. Due to variable intracranial electrode coverage across our five subjects, we identified variability in regions identified as effective biomarkers. In addition, we identified variability in model performance across subjects, and we hypothesize this variability is due to the quantity and range of pain reports we were able to gather. This work in identifying subject-specific biomarkers of pain can help broaden our understanding of pain mechanisms from an electrophysiology perspective and inform future pain management therapies by providing real-time estimation of a patient's pain levels.

5.2 Methods

Subjects

Collection of human intracranial data is detailed in Chapter 2: General Methods. Data from five subjects were analyzed. Subject demographics can be found in Table 5.1.

Neural data acquisition

Clinical data were collected as described in Chapter 2: General Methods. Data spanned 5-7 days, and visualizations of subject electrode placement can be found in Figure 5.1.

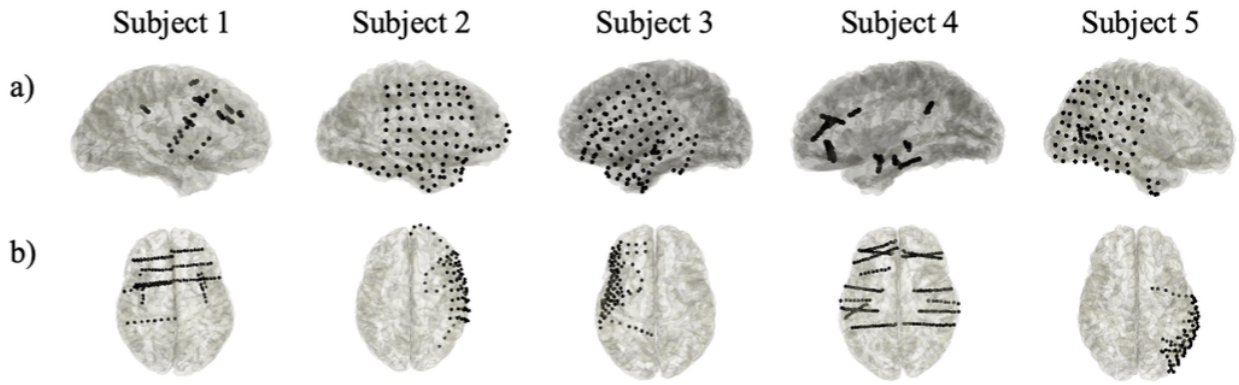


Figure 5.1: Electrode coverage across five subjects. Subjects 1, 4 had stereo-electroencephalography (sEEG) electrodes, and subjects 2, 3, 5 had electrocorticography (ECoG) electrodes.

Visual analog scale pain reports

As part of routine clinical care, subjects are asked to report a verbal pain score from 0-10 based on a visual analog scale (VAS). Subjects are typically asked to report their pain several times a day throughout their week-long stay at the hospital, and the number of total pain reports for each subject is reported in Table 5.1. Distributions of subject pain reports are provided in Figure 5.3a.

To define a high versus low pain state, we calculated the median pain score for each subject and used the median value as the dividing threshold. This method had been used previously[85] and promoted even group sizes for a binary classification task.

Table 5.1: Subject demographics.

Subject	Age	Sex	Channels	VAS Scores	Type
1	24	F	129	27	sEEG, bilateral
2	33	F	93	37	ECoG, right hemisphere
3	43	M	113	56	ECoG, left hemisphere
4	32	M	200	37	sEEG, bilateral
5	20	M	87	60	ECoG, right hemisphere

Data processing

The data processing pipeline is presented in Figure 5.2. After time-aligning neural data with the VAS reports, five-minute windows of neural data centered around the time of the pain report was extracted. Our primary assumption here was that the subject's pain experience did not change significantly during the five-minute window. Each window was further segmented into 30-second time segments for data augmentation.

Neural data were preprocessed using notch filters at 60 Hz and harmonics, and we applied a 5th order 200 Hz Butterworth low pass filter to remove high-frequency noise. Channels were visually inspected and removed if excessive noise or artifact remained present.

Neural features calculated included power-in-band, Spearman's correlation coefficient, and coherence. Power-in-band of canonical frequency bands (delta (0.1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), gamma (30–55 Hz), and high gamma (65–200 Hz)) were calculated by summing the square of the analytical signal, which was obtained by bandpass filtering and performing a Hilbert transform. Correlation coefficient was calculated in the time domain between all channel pairs. Coherence was calculated between all channel pairs for each canonical frequency band.

Model selection and training

For all classification models, we trained on 70% of the data and evaluated model performance on the held-out 30%.

Single-channel classification: Each single-channel feature was used to train an independent model to predict subject pain levels. This process allowed us to identify specific channel and frequency band features that were most predictive of subject pain levels. For predicting binarized high vs. low pain levels, we fit a logistic regression model and evaluated model accuracy on the held-out dataset. For predicting discrete VAS reports, we fit a linear regression model and reported R-squared values.

Multi-channel classification: The top 10 performing channels identified in the single-channel binary classification task were used to construct a pain network, where correlation and

coherence features were calculated across the 10 identified channels. All features—individual channel power-in-band, and multi-channel correlation and coherence features—where used as input for model training. Models used for multi-channel classification included logistic regression, decision tree, random forest, and linear discriminant analysis. Model accuracies were reported for both binary and discrete classification tasks.

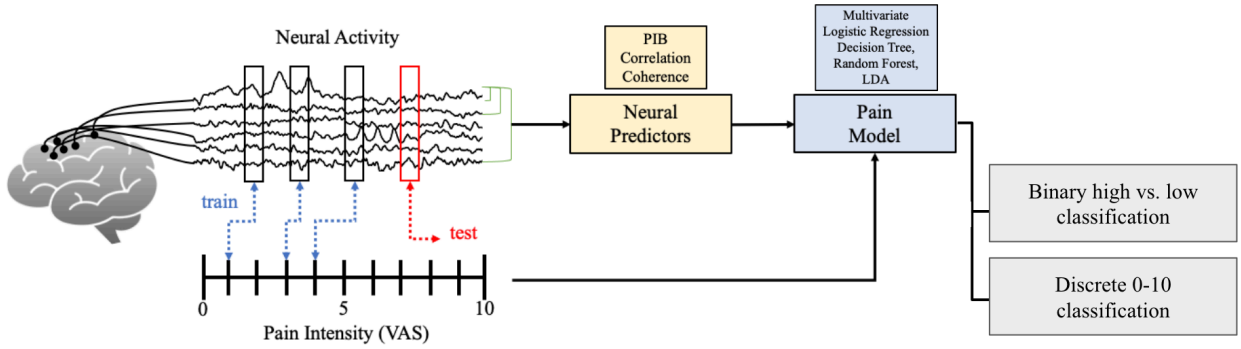


Figure 5.2: Overview of pain model pipeline. Neural activity was recorded using intracranial electroencephalography (iEEG) in epilepsy patients. Five-minute windows of neural activity centered around VAS pain reports were extracted and used to calculate neural features such as power-in-band (PIB) and connectivity metrics. These input features were used to train a pain classification model using a set of model architectures. There were two classification goals: predicting binary high vs. low pain state and predicting discrete VAS scores.

5.3 Results

Best performing pain classification features vary between subjects

All channel and power-in-band combinations were evaluated across subjects to identify best performing channel and frequency band pair, and results are displayed in Figure 5.3. For the binary classification task, top performing features across subjects ranged from a 68% prediction accuracy in Subject 5 to a 99% prediction accuracy in Subject 2, where chance prediction was 50%. Beta, gamma, and high-gamma bands were included in the top performing features, while regions varied.

Top performing features for fitting a linear model overall performed poorly across subjects, with R-squared values of 0.15 – 0.27 across four of the five subjects and 0.71 in Subject 2.

Except in Subject 5, none of the subjects shared the same region between the binary classification task and the linear model fit.

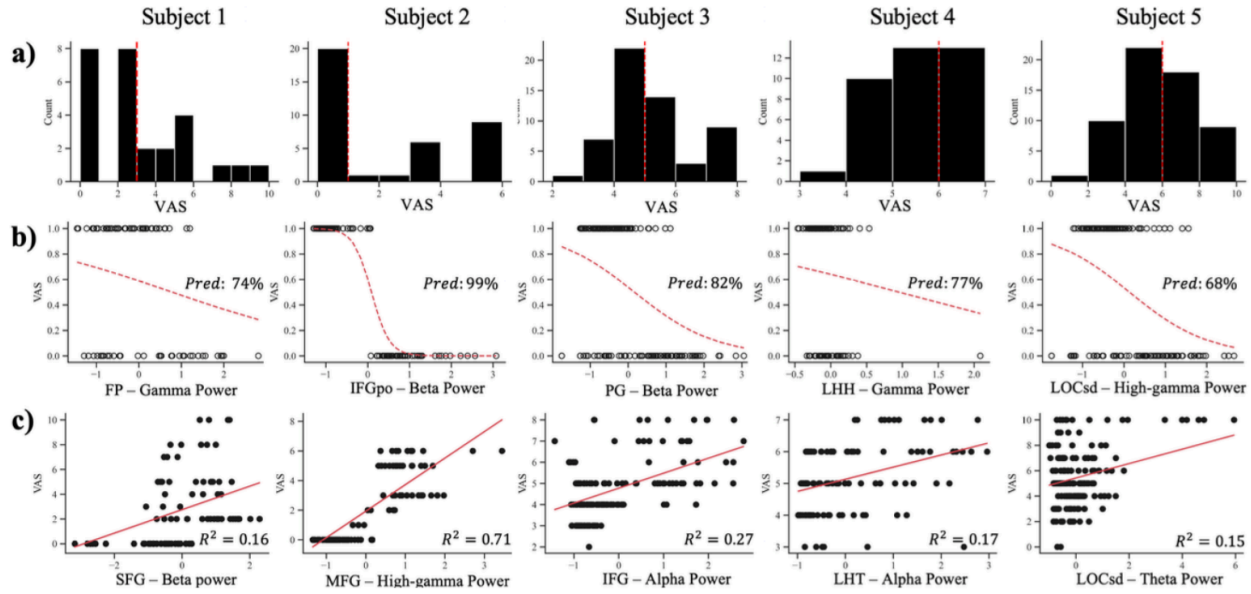


Figure 5.3: Best performing single-channel pain classification models across subjects. (a) Distribution of pain scores across the five subjects. The red dotted line indicates median pain score, which we used to define high and low pain states. (b) Best performing single channel and frequency band for binary low vs. high pain classification task with prediction accuracy reported. (c) Best performing single channel and frequency band in fitting a linear model to reported pain scores. FP – frontal pole; IFGpo – inferior frontal gyrus pars opercularis; PG – precentral gyrus; LHH – left hippocampal head; LOCsd – lateral occipital cortex superior division; SFG – superior frontal gyrus, MFG – middle frontal gyrus; IFG – inferior frontal gyrus; LHT – left hippocampal tail.

Random forest model outperformed in multivariate prediction

The top 10 performing channels in the univariate binary classification task was used to create a pain network, where only features within the network were included in model training and evaluation. Supplemental Table 5.1 includes all channels included in subject-specific pain networks. Tables 5.2 and 5.3 summarize the results across subjects and models tested in binary and discrete classification tasks respectively. In both binary and discrete classification, the random forest model outperformed all other models in all subjects, except in the discrete classification task in Subject 4, where all models performed poorly in comparison to the other

subjects. Averaging across subjects, the random forest model had prediction accuracies of 80% and 65% in the binary and discrete classification tasks respectively.

We observed that inter-subject variability of model accuracies was high in comparison to inter-model variability. Notably, Subject 2 had the highest model performance across all classification tasks, and we hypothesize that the bimodal shape of their pain report distribution contributed to the model's ability to learn effective features for pain prediction. In contrast, Subject 4 tended to have the lowest model performance, in addition to having the smallest range of pain reports (VAS~ [3, 6]).

Table 5.2: Multi-channel binary VAS prediction accuracies across models and subjects.

Subject	Logistic Regression	Decision Tree	Random Forest	Linear Discriminant Analysis
1	0.68	0.72	0.75	0.58
2	0.95	0.95	1.0	1.0
3	0.77	0.74	0.79	0.56
4	0.70	0.60	0.72	0.61
5	0.71	0.69	0.76	0.57
Mean	0.76	0.74	0.80	0.66

Table 5.3: Multi-channel discrete VAS prediction accuracies across models and subjects.

Subject	Logistic Regression	Decision Tree	Random Forest	Linear Discriminant Analysis
1	0.56	0.59	0.70	0.41
2	0.90	0.86	0.95	0.80
3	0.57	0.58	0.68	0.20
4	0.37	0.39	0.37	0.33
5	0.50	0.49	0.57	0.23
Mean	0.58	0.58	0.65	0.39

5.4 Discussion

In this chapter, we identified subject-specific pain biomarkers using both single channel features and multi-channel features across five subjects. With single feature logistic regression models, we identified specific regions and frequency bands that were capable of distinguishing between high versus low pain states at 68% - 99% percent accuracy. Using multi-channel features, we additionally compared several models and found that a random forest model was most effective at predicting both high vs. low pain state and predicting discrete VAS values across all subjects. These results contribute towards building patient-specific models of pain, which can help us better understand the diversity of pain experiences in addition to informing biofeedback therapies such as adaptive neurostimulation.

We demonstrated using both single-channel and multi-channel models, where multi-channel models included features from multiple regions. Due to the known multiregional changes that occur with pain, we hypothesized that models that included multiregional features performed better than single-channel models; however, we observed that both model types performed similarly in classification tasks. One explanation is that the relevant multi-channel features may have been highly correlated with the best performing single-channel feature, which would not have improved model performance. We believe that a larger dataset with more observations of pain reports would help evaluate the efficacy of using multi-channel models over single-channel models in predicting subject pain.

Our pain predictors were likely representative of acute head pain that our subjects were experiencing post-surgery. Further work is needed to determine whether these subject-specific biomarkers can be extended to chronic pain, which is of particular interest in adaptive neurostimulation applications.

We identified variability in model performance across subjects, and we hypothesize that the quantity and variability of pain reports are primary contributors to model performance. Our data is limited to the seven-day stay of these epilepsy patients, where clinicians gather patient pain reports anywhere from 3-8 times a day. We expect that our models would have improved prediction accuracies with data over a longer period. In addition, we observed that the distribution of pain reports varied among subjects, where some subjects utilized the full

range from 0-10, while others had a more limited range. This difference could be due to differences in how patients report pain or simply that some patients had more monotone experiences of pain. As we observed that our models performed worse in subjects with a smaller reported distribution of pain, there may be strategies to improve the pain labels, such as using more comprehensive pain reports[90]. Ultimately, our models rely on there being meaningful variability in patient experiences and reports of pain in order to identify pain-related biomarkers.

Pain was categorized as low vs. high using a median value threshold, but in practice, a clinical assessment would be needed to identify an appropriate threshold for the subject. Determining this pain threshold would be a necessary step towards building a feedback-based pain management strategy, and there has already been interest in using neurostimulation applications for chronic pain, including in spinal cord stimulation (SCS) and deep brain stimulation (DBS). Incorporating a real-time pain biomarker could improve current stimulation therapies by administering therapy only when needed, preventing overtreatment, and extending device battery life.

Careful considerations must be made when incorporating these real-time biomarkers in standard clinical care, especially when relying on biomarkers to know when to administer pain treatment. We currently lack information about the risk for addiction to these stimulation-based treatments and should be wary especially due to the high risk of addiction from current pain medications. Finally, when considering the use of automated devices that administer therapy only when needed, we must understand how the patient interacts with the device and whether there are better clinical outcomes when the device is fully automatic versus if the patient also has the option to control their treatment.

5.5 Supplemental Figures

Supplemental Table 5.1 Top ten performing channel-frequency bin models for pain network selection.

Subject	Region	Channel	Band	Logistic Regression Accuracy
1	Middle frontal gyrus	MFG15	high-gamma	0.741
	Paracingulate gyrus	PG23	high-gamma	0.741
	Frontal pole	FP7	gamma	0.741
	Cingulate gyrus	CGad4	alpha	0.728
	Postcentral gyrus	PG11	high-gamma	0.728
	Precentral gyrus	PG6	high-gamma	0.728
	Middle temporal gyrus	MTGpd9	gamma	0.716
	Cingulate gyrus	CGad3	gamma	0.716
	Postcentral gyrus	PG12	beta	0.716
	Frontal pole	FP13	high-gamma	0.704
2	Inferior frontal gyrus	IFGpo1	beta	0.991
	Temporal pole	TP10	beta	0.973
	Precentral gyrus	PG11	gamma	0.973
	Middle frontal gyrus	MFG9	beta	0.973
	Inferior frontal gyrus	IFGpt2	gamma	0.973
	Temporal pole	TP11	beta	0.964
	Frontal orbital cortex	FOC2	beta	0.964
	Supramarginal gyrus	SGad2	beta	0.955
	Central opercular cortex	COC1	gamma	0.955
	Precentral gyrus	PG1	beta	0.955
3	Precentral gyrus	PG7	beta	0.815
	Middle frontal gyrus	MFG7	alpha	0.804
	Middle frontal gyrus	MFG15	theta	0.792
	Superior temporal gyrus	STG6	high-gamma	0.792
	Inferior frontal gyrus	IFG7	beta	0.792
	Inferior frontal gyrus	IFG6	theta	0.792
	Inferior frontal gyrus	IFG5	alpha	0.792

	Middle frontal gyrus	MFG14	theta	0.786
	Superior temporal gyrus	STG1	high-gamma	0.786
	Middle frontal gyrus	MFG2	gamma	0.786
4	Hippocampus	RHH8	delta	0.775
	Occipital frontal	ROF6	high-gamma	0.775
	Hippocampus	LHH6	gamma	0.775
	Posterior cingulate cortex	RPCC16	high-gamma	0.766
	Hippocampus	RHH6	delta	0.766
	Occipital frontal	LOF6	beta	0.766
	Amygdala	RA12	theta	0.757
	Amygdala	RA7	alpha	0.757
	Anterior cingulate cortex	RACC8	alpha	0.757
	Anterior cingulate cortex	LACCS6	gamma	0.757
5	Occipital cortex	LOCsd2	high-gamma	0.683
	Occipital cortex	LOCsd6	theta	0.678
	Occipital cortex	LOCsd1	beta	0.667
	Supramarginal gyrus	SGpd2	theta	0.661
	Angular gyrus	AG4	alpha	0.656
	Angular gyrus	AG1	theta	0.656
	Occipital cortex	LOCid9	theta	0.639
	Occipital cortex	LOCid8	gamma	0.633
	Inferior temporal gyrus	ITGpd1	theta	0.633
	Inferior temporal gyrus	ITGpd4	gamma	0.628

6. Characterizing local field potential responses to patterned stimulation

The primary motivation for this work is to build an understanding of what measurable responses can be detected from stimulation. An electrical stimulation input will activate the population of neurons in the near vicinity and prompt an electrochemical response, and where we are able to record that response is largely correlated with what regions this neural population is connected to. Due to the interconnectedness and redundancy present in neural systems, in addition to ongoing neural activity, the response may not necessarily be obvious or easily measurable.

An added layer of complexity is that there is a vast parameter space of stimulation parameters that can be used for neuromodulation, and there are ongoing efforts in fine-tuning parameters or discovering new parameter combinations that will be more effective in stimulation-based therapies. While it is well established that different stimulation parameters produce different outcomes, there is less understanding or precedent on whether we can measure those differences.

Most neuromodulation work revolves around a target neurological condition or specific circuit, whereas this work takes a more generalized approach to neuromodulation. Our patient population tends to have either unilaterally or bilaterally distributed electrode implants, allowing us the unique opportunity to investigate a wider spatial range of stimulation responses across brain regions.

This chapter describes work that seeks to quantify measurable effects of stimulation using an experimental paradigm adapted from Yang et al.[36] An initial conference paper on this work was published in IEEE SMC 2022 and the following chapter is an extension of the work with updated methods.

[published IEEE SMC 2022] Sun S, Levinson LH, Paschall CJ, Herron JA, Weaver K, Hauptman J, Ko A, Ojemann J, Rao RPN; Neural responses to different stimulation conditions are separable in a low-dimensional subspace

[in preparation] Sun S, Levinson LH, Weaver K, Ko A, Grannan B, Hauptman J, Tsai J, Herron JA, Rao RPN, Ojemann J; A generalized approach of measuring neural response separability under varying stimulation inputs

6.1 Introduction

Electrical stimulation has emerged as a powerful tool for treating neurological conditions. By delivering targeted electrical pulses to specific regions, neural activity and circuitry can be modulated to promote symptom reduction. Specifically, deep brain stimulation (DBS) has become a well-established treatment for Parkinson's disease and essential tremor[91,92] and ongoing research seeks to expand DBS application to other treatment-resistant pathologies including in depression[93,94], obsessive compulsive disorder[95], bipolar disorder[96], and epilepsy[97].

Adaptive DBS is a rapidly growing area of research to improve and personalize stimulation therapies. Unlike traditional methods that use open-loop stimulation at fixed parameter settings, adaptive DBS adjusts stimulation settings based on real-time patient-specific needs– turning stimulation on and off and potentially even calibrating therapy dosage– for improved patient outcomes. To accomplish this, neural biomarkers that reflect symptom state can be monitored and leveraged in a closed-loop system. There have already been effective demonstrations of adaptive stimulation using neural biomarkers, where beta-power in the hand area of the primary motor cortex has been used to turn on and off stimulation for tremor management in essential tremor[4,5], and cortical gamma power and beta power in the subthalamic nucleus have been used to adjust stimulation amplitude for improved symptom management in Parkinson's disease[37,98]. Adaptive therapies have been shown to perform as well as open-loop stimulation while reducing side-effects and extending device battery life[4,37,38].

While significant advancements have been made in identifying neural biomarkers related to symptom state, methods for adjusting stimulation or selecting effective stimulation sites to modulate target biomarkers is less understood. In some cases, stimulation can be applied directly to the biomarker region to modulate its activity and reduce symptom state[98], but in other cases, the effective stimulation site is away from the biomarker site[4,5,9,37,39]. Pre-

stimulation activity[18,21,25] and resting state connectivity[20,22] have been used to help predict stimulation outcomes, but ultimately, we have poor understanding of how electrical stimulation affects neural electrophysiology in real-time and how the effects propagate to other regions. This lack of knowledge limits our ability to determine real-time optimal stimulation parameters and regions to achieve desired patient outcomes.

A growing body of work has been focused on modeling and predicting the real-time relationship between stimulation and neural response for adaptive therapies. Bolus et. al.[34] used optimal control theory to modulate stimulation intensity to control neural firing rates in rats, and Yang et. al[36] predicted stimulation-induced local field potential activity in monkeys using a state-space model and simulated closed-loop stimulation for alleviating mood symptoms. Translating these methods to humans requires much additional work. Another consideration is that the brain is a complex, dynamic system, and using the same stimulation input can often produce varying responses depending on current brain states[42,99]. Long-term electrical stimulation has also been shown to promote neural plasticity and adaptation[100], potentially requiring modified stimulation dosage to maintain effectiveness. These complex interactions have contributed to the challenge of building effective stimulation models.

To build a more comprehensive understanding of how stimulation interacts with the brain, one approach is to measure and map neural responses to different stimulation inputs over time. By characterizing stimulation responses, we gain crucial insights into (1) what stimulation parameters can modulate region activity and by how much, (2) what regions can be modulated with stimulation input, and (3) the stability of stimulation response over a period of time. By understanding the intricate relationship between stimulation and response, we strongly believe that these efforts will help improve and personalize future stimulation methods for adaptive closed-loop therapies.

Our research aim was to determine whether varying electrical stimulation inputs produced distinct, measurable changes in neural electrophysiology. Given the varied, subject-specific electrode coverage, our secondary aim was to develop a data-driven method of identifying regions sensitive to changes in stimulation that was agnostic to stimulation site, electrode coverage, and pre-existing knowledge of region-specific circuitry. Our paper contributes to efforts of characterizing the effects of stimulation interventions– specifically the spatial reach

of stimulation effect and the range of possible biomarker outcomes we expect from stimulation.

6.2 Methods

Subjects

Our subjects (n=9) comprise of patients with intractable epilepsy undergoing seizure monitoring at Harborview Medical Center in Seattle, Washington, USA. As part of routine monitoring procedures, subjects were implanted with multiple stereo-electroencephalography (sEEG) electrodes (0.86mm diameter, Ad-Tech Med Instr Corp, USA), and implant regions were determined entirely for clinical monitoring purposes. Electrode placement varied between subjects. Data collection was approved by the Institutional Review Board of the University of Washington, and all subjects gave written, informed consent. Subject demographics can be found in Table 6.1, and visualizations of individual electrode placement can be found in Supplemental Figure 1.

Table 6.1: Subject Demographics

	Age	Sex	Handedness	Implant hemisphere
S1	23	M	L	bilateral
S2	37	F	R	left
S4	25	M	R	bilateral
S5	36	M	R	left
S6	24	M	R	bilateral
S7	44	M	R	bilateral
S9	22	M	R	bilateral
S10	26	M	R	bilateral
S11	46	F	R	bilateral

Data Acquisition

Data were collected using the Tucker Davis Technologies (TDT) as described in Chapter 2: General methods.

Stimulation protocol

Stimulation waveform and electrode selection were performed as described in Chapter 2: General methods.

Stimulation experiment

Our stimulation input consisted of a repeating sequence of stimulation that varied in amplitude and frequency at fixed, 5-second intervals (Figure 6.1A). By default, a sequence comprised all combinations of stimulation frequencies [10,50,100 Hz] and amplitudes [2,4 mA] and included a no-stimulation condition. Each stimulation presentation was individually tested prior to the experiment and either removed if ictal activity was observed after stimulation presentation or the amplitude was decreased to sub-perceptual threshold in the case of subject perception of stimulation (Subject S2).

Stimulation conditions were presented in a random sequence at 5-second intervals, and this sequence was repeated over time, ranging from 4-14 total repetitions (See Table 6.2 for subject-specific details). These fixed-sequence repetitions allowed for measuring changes in stimulation response over time, controlling for potential neural drift and stimulation ordering effects.

A pre-experiment resting state (140-350 s) was performed to establish a baseline measurement. While not initially considered in our research design, we also analyzed the post-experiment resting state (4-135 s) for any residual effects from the stimulation experiment (Figure 6.1B).

Our stimulation experiment was adapted from Yang et. al[36], where researchers designed a repeating stimulation sequence to measure and predict stimulation response given only the stimulation parameters. Their experiment was performed in non-human primates, allowing them to run stimulation experiments for longer durations of time. Due to the nature of working in a human epilepsy patient population, it was necessary to modify the experiment such that the time-in-room duration did not exceed three hours, as defined in our IRB research protocol. Design decisions for the number of stimulation parameters and number of repetitions were largely influenced by this time constraint.

Table 6.2: Stimulation experiment details by subject, including number of recording channels after removing channels with poor signal quality, stimulation channels, stimulation parameters, and sequence repetitions.

STIMULATION EXPERIMENT DETAILS

	active recording electrodes	stimulation channel(s)	stimulation parameters	total stim params (+0)	total sequence repetitions
S1	109	RACC 7-8	[10, 50, 100] Hz at [2, 4] mA	6	5
S2	113	LTPJ 9-10	[50, 100] Hz at [0.5, 1, 1,5] mA	5	14
S4	113	LOF 7-8	[10, 50, 100] Hz at [2, 4] mA	7	4
S5	125	LOF 7-8	[10, 50, 100] Hz at [2, 4] mA	6	10
S6	125	RPCC 4-5	[10, 50] Hz at [2, 4] mA	5	10
S7	126	LPCC 11-12	[10, 50, 100] Hz at [2, 4] mA	7	10
S9	124	LOF 13-14, LOF 8-9	[50, 100] Hz at [1, 2, 4] mA	7	10
S10	124	LOF 3-4	[50, 100] Hz at [1, 2, 4] mA	7	10
S11	105	LSTG 5-6	[50, 100] Hz at [1, 2, 4] mA	7	10

Data processing

Preprocessing methods

Channels that had unrecoverable signals due to visually excessive noise or artifacts were removed from analysis. Data preprocessing was performed in the Python environment. Our data cleaning pipeline included a 100 Hz low-pass filter, down-sampling to 200 Hz for computational efficiency, a 1 Hz high-pass filter for removal of baseline drift, and a notch filter at 60 Hz to remove line noise. All filters used a zero-phase 2nd order Butterworth filter. For handling stimulation artifacts, we performed a comprehensive evaluation of artifact removal methods in Chapter 3 and selected notch filtering as it was the most aggressive filtering approach. Notch filters were applied at stimulation frequencies [10, 50, 100 Hz] and their harmonics, and these frequencies were removed from analysis. We observed additional artifacts time-locked to stimulation transitions, and in response we removed 500-1000 ms of signal at the beginning and end of each stimulation presentation.

Frequency analysis

Power spectral densities were calculated by using Welch's method[67] with a Hanning window. We calculated the time-frequency spectrogram with a Tukey window of 250 ms and 1/8th overlap (Figure 6.1C). The resulting spectrogram had a time resolution of approximately 217

ms and frequency intervals of 2 Hz. Power values were transformed into log space. We additionally removed all frequency bins that overlapped with stimulation frequencies and their harmonics.

We performed outlier removal by identifying power values 10 times greater than the interquartile range of each frequency interval and setting it to the median value. Baseline correction was performed by calculating the average log power during the pre-experiment resting period and subtracting the resting state power from experimental power values.

Dimensionality reduction

Our resulting dataset was spectrogram data across all recording channels (time x frequency x channels). For each channel, we performed dimensionality reduction via principal component analysis (PCA), compressing the frequency space (dimensions = 41). We note that other papers perform dimensionality reduction across channels, while selecting a single frequency band of interest[101]. Our method allowed us to leverage using multiple frequency band features at the computational cost of performing this task across all channels. Using a grid search method during hyperparameter testing, we selected a data dimensionality of 10 to use for our prediction tasks. When splitting the data into training and validation sets, we fitted the PCA model on the training set only to ensure that there was no biasing from the validation set.

Supervised learning approach

Our primary task involved measuring the separability between stimulation responses. We used a support vector machine (SVM) model and evaluated its performance on a validation dataset using k-fold cross validation. The folds were defined as individual 5-second stimulation presentations, such that the training set did not include any data from validation epochs to avoid bias[102]. The number of folds was equal to the number of stimulation repetitions performed, which differed among subjects.

We performed permutation testing on scrambled labels to generate a null hypothesis distribution and measure the statistical significance of our test accuracies. Permutation testing is a non-parametric method to control for family-wise error rate[103].

Group-level vs. pairwise classification

We performed two types of classification tasks: group-level and pairwise classification (Figure 6.1D). Group-level or one-vs-all classification is a multiclass classification task, where a support vector learns to separate a class from all the other classes, and a complete model includes a support vector for each class. Pairwise classification, or one-vs-one classification, is a binary classification task between a pair of stimulation responses. We performed pairwise classification for all combinations of stimulation pairs to obtain a more granular understanding of which stimulation responses are more separable than others.

Separability vs. sensitivity

We define separability as the model accuracy of a trained supervised model in predicting distinct responses to distinct stimulation inputs. Separability can be measured among a group of stimulation inputs and between pairs of stimulation inputs. While group-level separability effectively describes how distinct each stimulation response was with respect to all other responses, pairwise separability only compares pairs of stimulation responses, allowing a more granular description of which pairs are more easily separable than others. To summarize pairwise separability as a single metric, we define sensitivity as the percentage of separable stimulation pairs in a given region. Our sensitivity metric tells us how many combinations of stimulation input pairs produce distinct responses. In this chapter, we use separability in a group-level descriptor and sensitivity as a pairwise descriptor. Both metrics can inform us of regions that could be good candidates for capturing distinct stimulation responses.

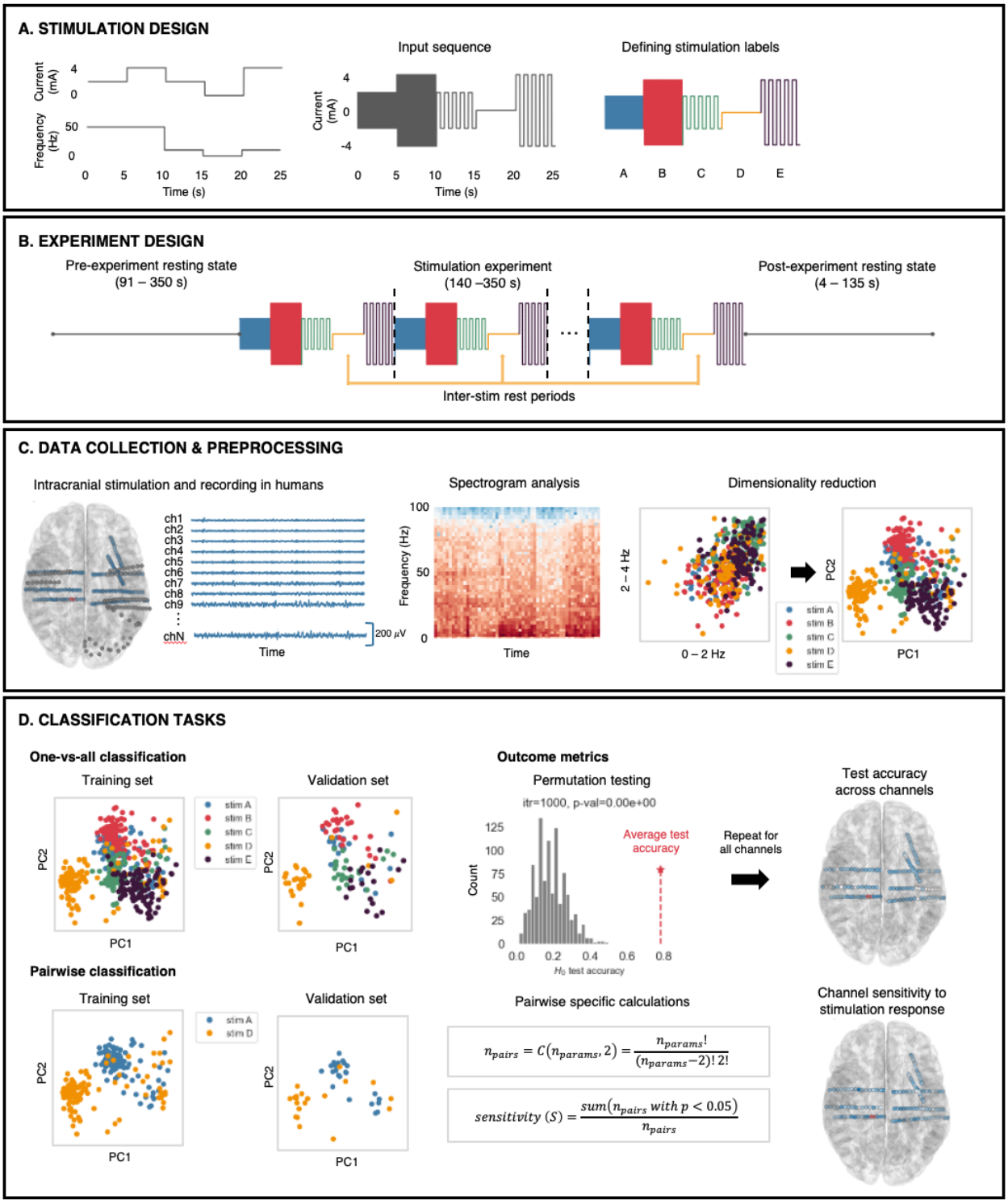


Figure 6.1: Overview of stimulation experiment and data analysis methods. A: Electrical stimulation design involved selecting from a set of stimulation amplitudes and frequencies, including a no-stim condition, to create a randomized input sequence. Each stimulation

condition had a fixed duration of 5-seconds. B: Experimental design included a pre-experiment resting state to record baseline activity, the stimulation experiment which involved repeating sequences of stimulation, and a post-experiment resting state to measure any potential experiment-induced effects. C: Data were collected from up to 128 electrodes from custom patient montages. After signal preprocessing and cleaning (not shown), time-frequency spectrograms were generated for each channel. Each stimulation response could be mapped to an N-dimensional frequency space, and dimensionality reduction was performed to compress the frequency dimension. D: For each channel, one-vs-all and pairwise classification was performed using support vector machine (SVM) models to evaluate the separability of stimulation response to differing inputs. Permutation testing was used to evaluate model performance, and the accuracies were mapped to patient electrodes for visualization. Specific to pairwise classification, each stimulation pair combination trained an SVM model, and a sensitivity metric was defined as a summary metric across all stimulation pairs, which was also mapped to patient electrodes.

6.3 Results

Identification of channels with separability better than chance in all subjects
In all subjects, we identified a set of recording channels that performed better than chance at separating stimulation conditions with a one-vs-all SVM classifier (Figure 6.2A). Chance performance was defined with permutation testing and depended on the number of stimulation conditions presented per subject. The percentage of channels that had better-than-change performance ranged between subjects, and we hypothesize that this wide range is largely dependent on both the stimulation channels used in addition to the patient-specific location of the recording channels.

While the percentage of channels performing better than chance varied across subjects, the accuracy values themselves ranged from 28% to 60% across the best performing channels across subjects, noting that the values for chance accuracy and p-value cutoff changed depending on the number of stimulation parameters.

When examining individual channel performance for subject S6, we discovered that top performing channels created visually separable clusters of stimulation response, while the majority of statistically significant channels seemed to only pull out a single stimulation condition, in this case, the inter-stim rest condition (Figure 6.2B). Furthermore, we observed that the first PC in channel LANT4 seemed to pull out the variance possible for a binary stim vs. rest classification task, indicating that the largest contribution to the variance for this channel comes from the differing responses between stimulation and rest. These results help define that better-than-chance performance could vary from being successful at a rest versus stim task, to being successful at separating all stimulation conditions. Additional plots of the top performing channels for each subject can be found in Supplemental Figure 2.

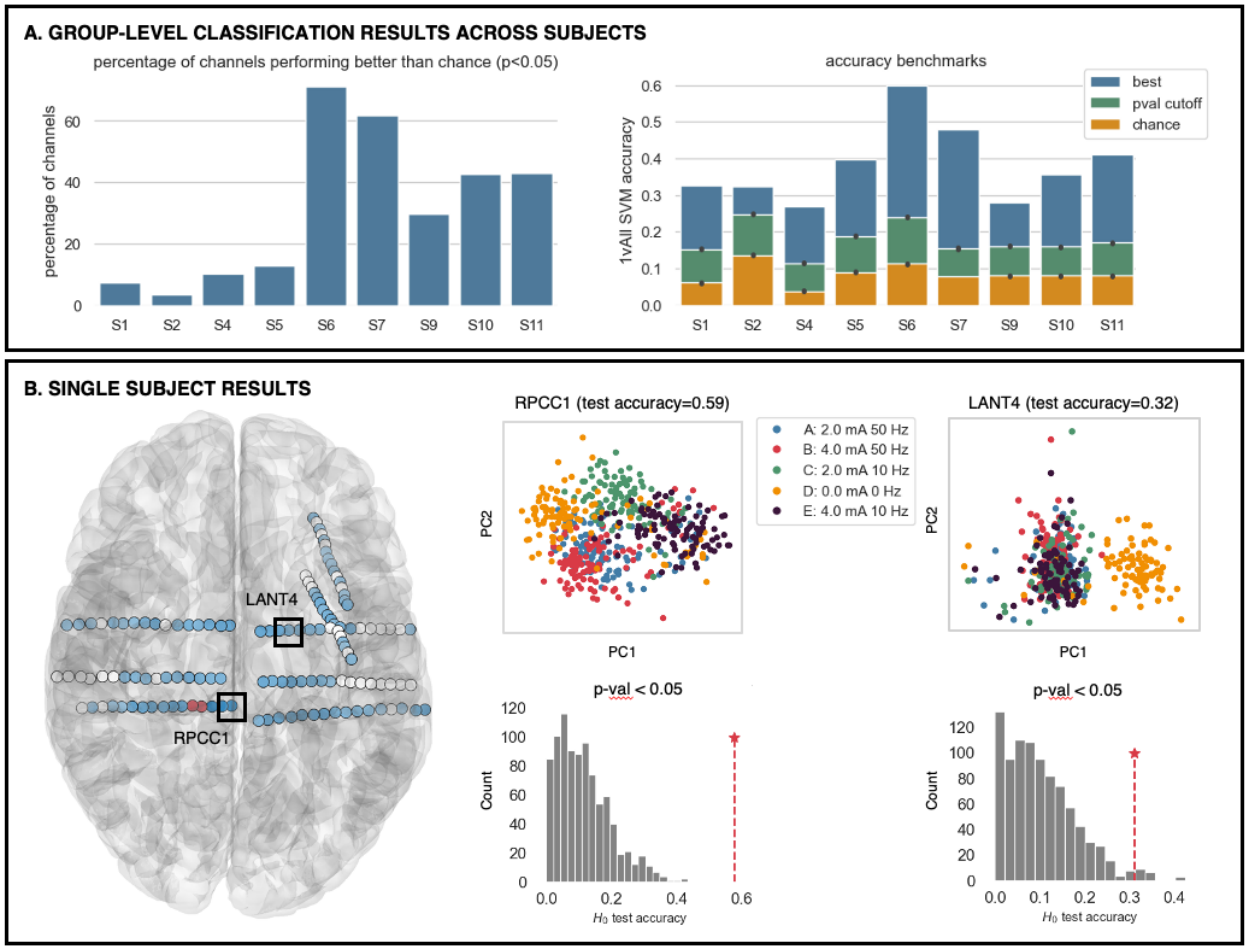


Figure 6.2: Group-level separability results. A: Subjects ranged in percentage of channel performing better than chance at a group-level classification task, which we hypothesize is related to stimulation channel selection and subject-specific channel placement (left). Accuracy benchmarks aggregated across channels, including the best performing channel, the p-value cutoff at the 95th percentile, and the average chance accuracy is shown (right). B: Example separability results for a single subject (S6), highlighting two channels with varying separability accuracies. While both channels reported accuracies significantly greater than chance, channel RPCC1 had improved performance, with a test accuracy of 59% and visibly separable stimulation clusters, whereas channel LANT4 had a test accuracy of 32%, where only stimulation condition D (inter-stim rest condition) was separable.

Pairwise prediction revealed stimulation inputs more separable than others

Figure 6.3 summarizes pairwise classification results for subject S5, demonstrating that some stimulation inputs were separable while others were not. For channel LOF6, two particular

pairwise classification results are highlighted: stimulation condition A vs. D (68% model accuracy) and stimulation condition C vs. D (95% model accuracy). This particular set of comparisons is exemplary in demonstrating the benefits of pairwise classification as it provides additional information of which specific stimulation pairs are separable compared to group-level classification, where all stimulation conditions are considered as a whole.

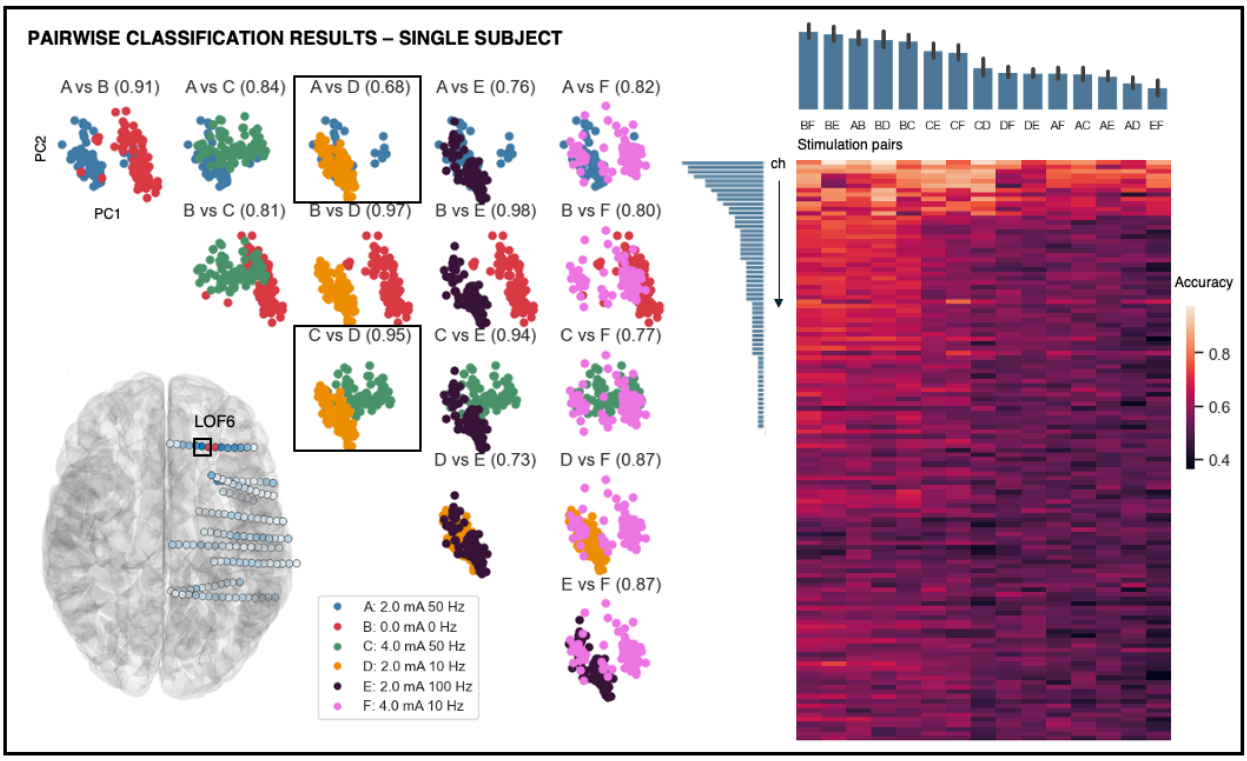


Figure 6.3: Pairwise classification results in subject S5. Visualization of stimulation response pairs (left) in an example channel LOF6 revealed that some stimulation responses were more easily separable than others. Here, stimulation A vs. D (68% model accuracy) had relatively poor separability in comparison to stimulation C vs. D (95% model accuracy). A sorted heat map visualization of pairwise classification accuracy aggregated across all recording channels (right). The above bar plot summarizes the average separability for each stimulation pair across channels, giving a ranking of most to least separable stimulation pairs. The left bar plot represents channel sensitivity, or the percentage of stimulation pairs with SVM classification accuracies at the 95% significance level defined with permutation testing.

Another observation is that the inter-stim rest condition (stimulation condition B, red), had high separability from the rest of the active stimulation conditions. When ranking stimulation

pairs from most to least separable across channels, we observed that the top five stimulation pairs included this inter-stim rest period. This result indicates that out of all the stimulation conditions, the inter-stim rest condition is the most easily distinguished when compared to any of the other stimulation conditions. Similar findings were observed in many of the other subjects, and additional plots of stimulation pair rankings for each subject can be found in Supplemental Figure 3.

In addition to aggregating results across all channels, we calculated and ranked each channel's sensitivity metric, which summarizes the percentage of stimulation pairs that were separable at the 95% significance level. The best performing channels may not necessarily be able to separate all pairs of stimulation input, but for the pairs that are separable, they are ideal candidates for distinguishing between stimulation inputs.

Separability and sensitivity metrics are highly correlated across subjects

Here, we characterized the relationship between group-level separability and pairwise classification results (Figure 6.4A). In example subject S7, we observed a strong positive correlation between channel separability and sensitivity, with a Pearson correlation coefficient value of 0.9115. Intuitively this result makes sense, where channels that have high group-level separability results also will have high pairwise separability, leading to high sensitivity. We observed the same trend across all subjects (Supplemental Figure 4). The reported p-values test whether the slope is statistically significant from zero, which is the case for subject S7 and for all other subjects.

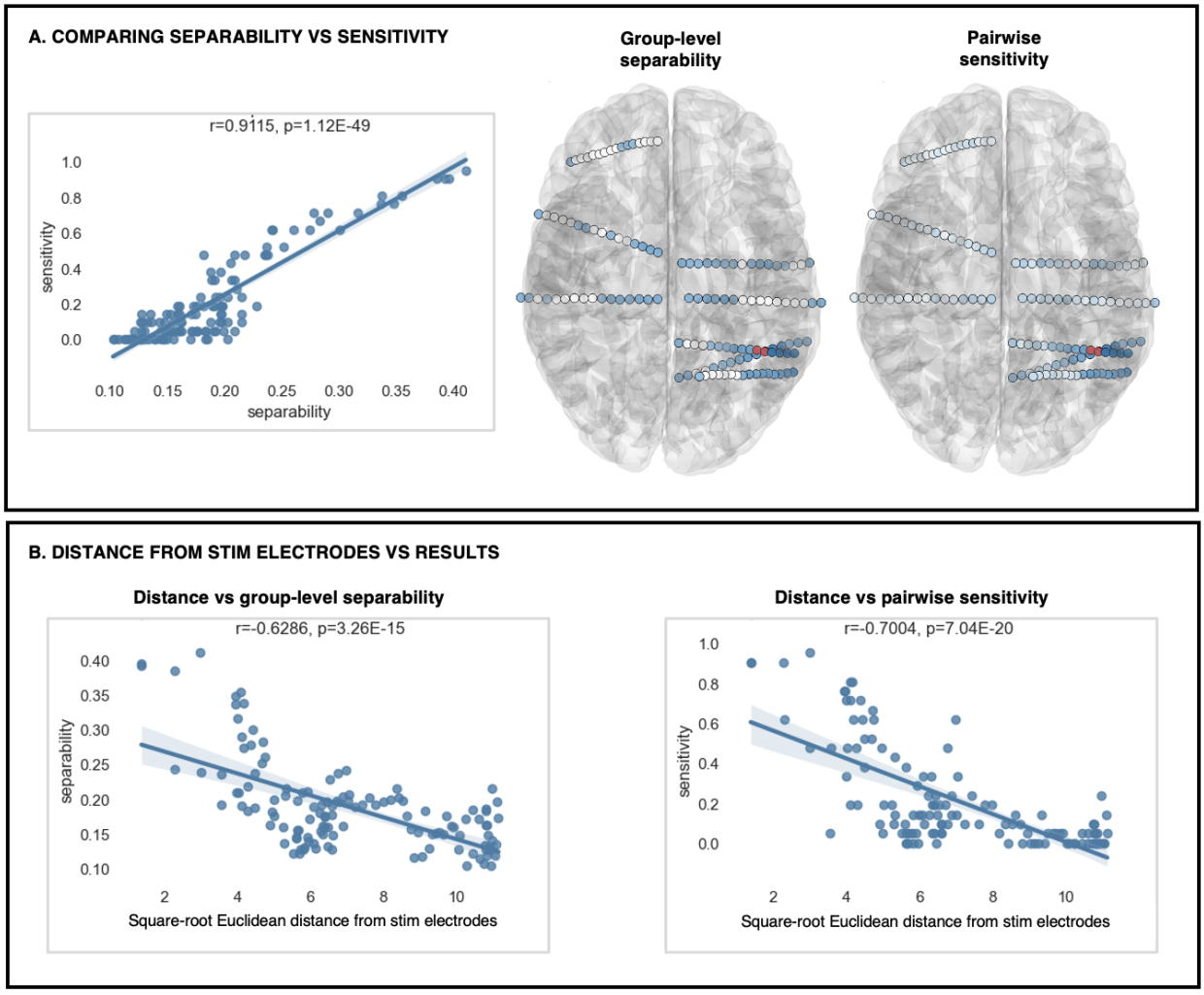


Figure 6.4: Relationships between group-level separability, pairwise sensitivity, and distance from stimulation electrodes using a linear regression fit in example subject S7. A: Subject S7 reported a high correlation (Pearson's correlation coefficient = 0.9115) between separability and sensitivity across channels. The reported p-value indicated that the slope is statistically significant from zero. A visualization of electrodes projected onto an MNI brain displays color-normalized values of group-level separability accuracy and pairwise sensitivity respectively. B: An inverse correlation between our results and the square-root of electrode distance from stimulation electrodes was observed, indicating that electrodes closer to the stimulation region tended to perform better in predicting stimulation response.

Distance from stimulation electrode inversely correlated with separability and sensitivity

We additionally evaluated whether distance from the stimulation channels was correlated with channel separability and sensitivity. We calculated the Euclidean distance from each channel to a point between the stimulation channels and initially observed an inverse exponential relationship between distance and our metrics ($1/r^2$), which is similar to the relationship between distance and spatial charge spread. As a result, we instead used the square root distance to fit linear regression models with our metrics (Figure 6.4B). Both separability and sensitivity metrics were inversely correlated with channel distance from stimulation electrodes, indicating that channels closer to the stimulation site tended towards higher separability and sensitivity values.

Stimulation amplitude positively correlated with pairwise separability

To better understand why some stimulation pairs were more separable than others, we investigated the potential relationship between pairwise classification accuracy and stimulation parameters (Figure 6.5). In example subject S5, we identified that stimulation amplitude had a positive linear relationship with classification accuracy, with a slope that was statistically significant from 0 at the 95% confidence level. Stimulation frequency and total charge delivered did not have a statistically significant correlation with pairwise classification accuracy. These results indicate that stimulation amplitude had the most impact in producing separable stimulation responses. Results aggregated across all subjects can be found in Supplemental Figure 5.

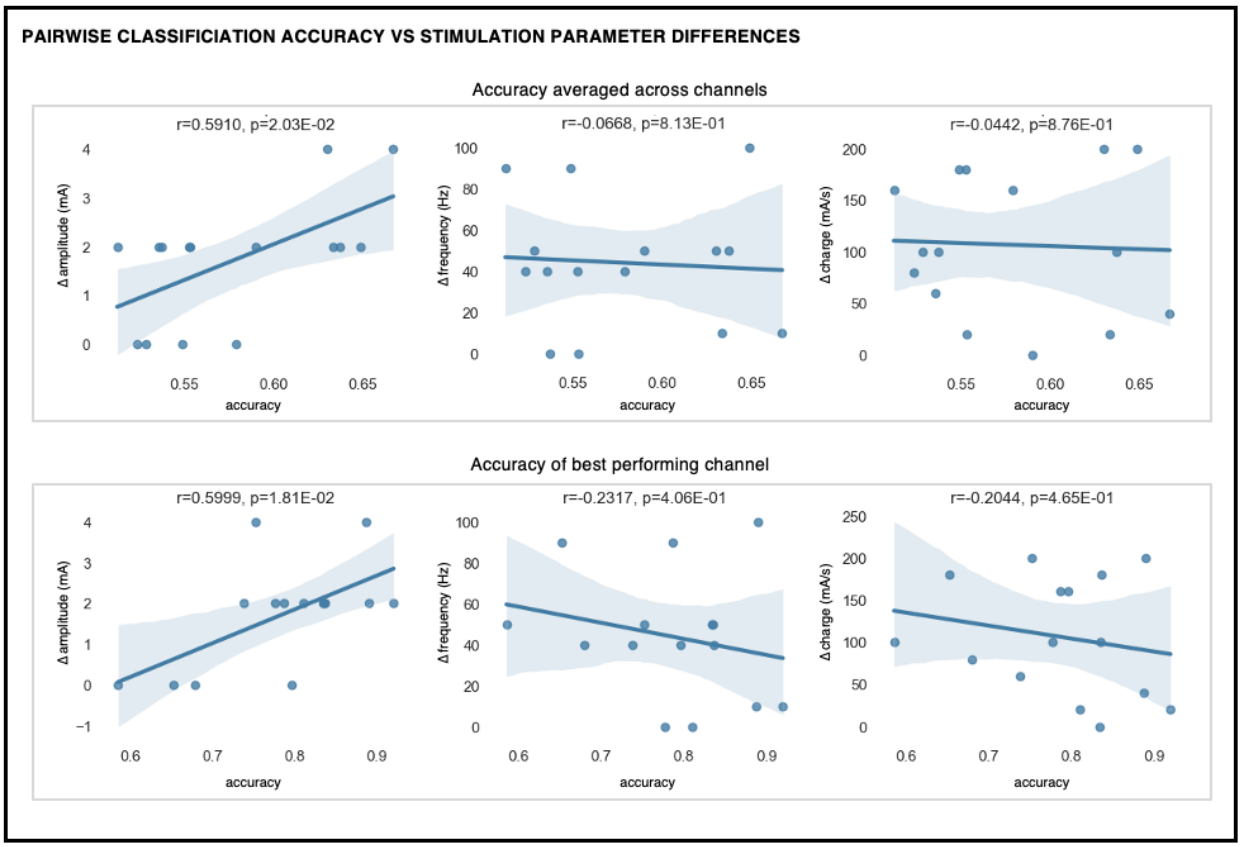


Figure 6.5: Linear relationship between stimulation parameters and pairwise classification results in example subject S5. Pairwise classification accuracies averaged across channels (top) or of best performing channel (bottom) were used to model linear relationships with stimulation amplitude (left), frequency (center), and charge delivered (right). Stimulation amplitude had a positive correlation with accuracy and was the only stimulation parameter that had a slope parameter that was statistically significant from 0 at the 95% significance level. Stimulation frequency and total charge delivered did not have a statistically significant relationship with pairwise accuracy.

Frequency response to stimulation remained stable throughout experiment

Given the novel patterned stimulation protocol, in addition to each subject having potentially different stimulation parameters and sequences, we evaluated whether there were changes in the response to stimulation over time, due to potentially several sources including stimulation adaptation or compounding effects from sequenced stimulation. For the duration of the experiment, we observed little to no change in stimulation response, specifically that the frequency response had little change across stim repetitions (Figure 6.6). This result indicates

that the response to stimulation remained stable throughout the duration of our experiment, which allowed us to perform our classification tasks under the assumption that responses were stationary and did not dynamically change over time.

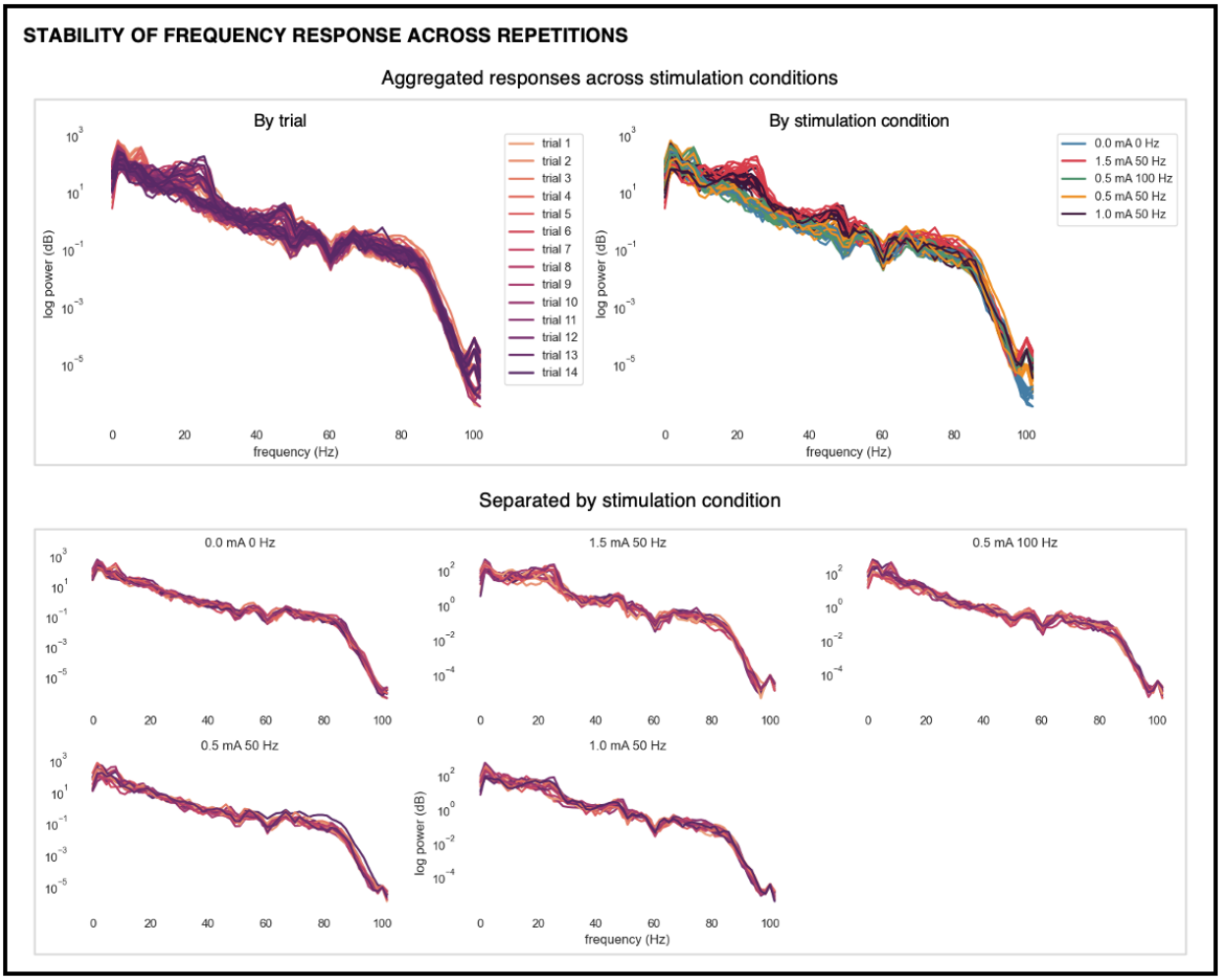


Figure 6.6: Frequency response to stimulation remained stable throughout experiment in example subject S2. Top: PSD plots displaying frequency response across all trials and stimulation conditions, in aggregate (left) and color-coded by stimulation condition (right). Bottom: PSD plots separated by stimulation condition demonstrated trial stability across all stimulation conditions.

6.4 Discussion

This chapter presented the results of characterizing human electrophysiology response to varying stimulation conditions. By employing an experimental paradigm involving repeating sequences of varying stimulation conditions, we measured responses to different stimulation inputs over time. We identified stimulation conditions that produced separable responses, in addition to identifying channels that were most effective in detecting these separable responses.

While overall group-level classification performance was low, we identified key channels, primarily those adjacent to stimulation channels that were capable of distinguishing between most stimulation inputs. More channels were capable of distinguishing between stimulation and rest but could not tell apart the different stimulation conditions. These results indicated that there was overlap in responses between stimulation inputs, meaning that not all stimulation inputs produced measurable, separable responses. Pairwise classification produced further evidence that only some stimulation pairs were separable, where we identified cases of both overlapping and non-overlapping stimulation responses.

Differences in stimulation amplitudes seemed to have the largest effect on separability. We identified a strong positive correlation between stimulation amplitude and pairwise classification accuracy, while stimulation frequency and total charge delivered seemed to have a slight negative correlation but no statistical significance. In other words, it's more likely for stimulation pairs to be separable if they have a large difference in stimulation amplitude, and it's unclear if a similar relationship exists for stimulation frequency and total charge delivered. In the case of non-separable stimulation pairs, it's possible that the difference in stimulation amplitude was not sufficient to produce distinct responses. One could re-attempt this experiment using a larger range of stimulation parameters, while also balancing that low amplitude stimulation runs the risk of producing smaller, harder to detect responses. Another consideration is that there is likely a limit to how much a region can be perturbed due to anatomical and biological constraints. While performing a comprehensive parameter search was not in the scope of this project, future work involves investigating the minimum required parameter alteration needed to produce a separable response. These

alterations could include stimulation amplitude, frequency, and potentially other stimulation parameters, or total charge delivered.

The variance of stimulation response contributed to response overlap, which increased the difficulty of our separability task. Having variance in response is expected, due to the ongoing activity that is separate from the stimulation experiment. Averaging the stimulation responses over a number of repetitions, we assumed that this underlying activity would not selectively confound stimulation responses and considered it as noise. We did not provide specific instructions for our subjects, and subject behaviors ranged from watching TV, reading on their phone, or chatting with someone in the room. These naturalistic behaviors can be useful for our experiment, where we can evaluate the stability of stimulation response on top of ongoing, undirected behaviors, but the drawback is that these behaviors may have introduced additional variance to our stimulation responses. It's possible that asking subjects to engage in a specific task throughout the experiment would reduce the response variance; however, it's unclear how the task may affect stimulation response. Current research efforts exist that aim to understand these dynamic interactions between electrical stimulation interventions and ongoing neural behavior[35,36].

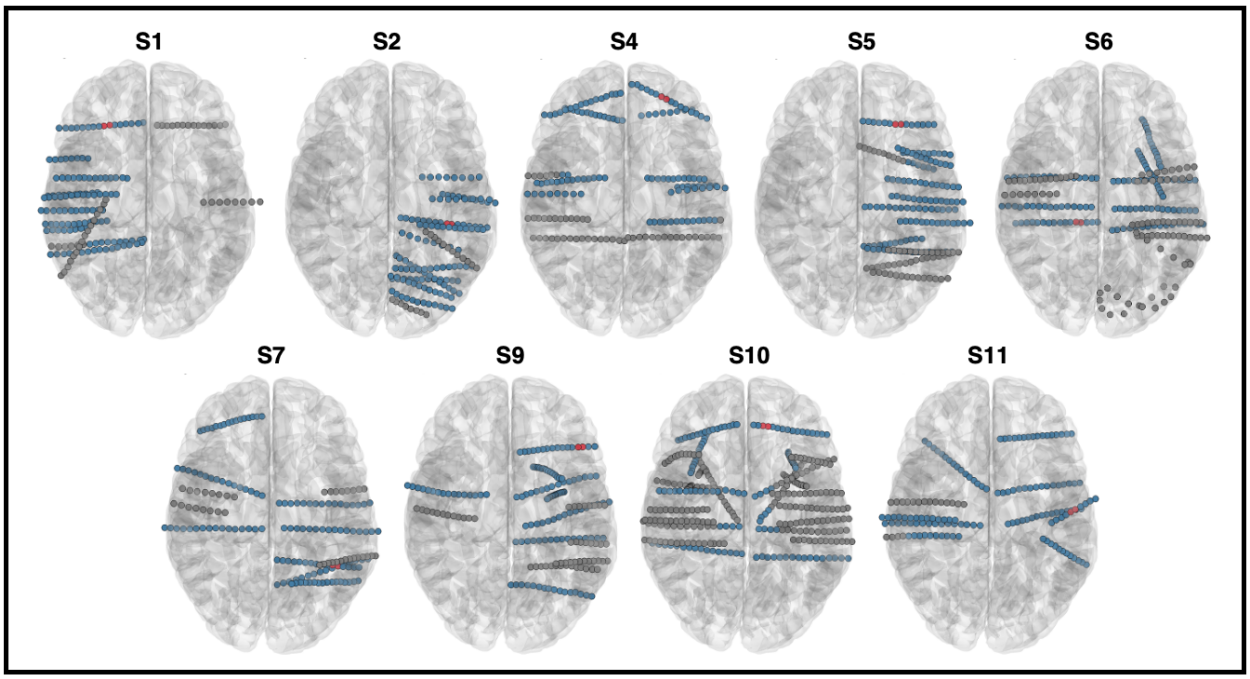
Finally, due to the nature of working with our human patient population, our total research time was limited such that our experiment length and number of repetitions may not have been sufficient to reduce the level of variance and effectively capture stimulation response.

Channels that were able to detect separable stimulation responses tended to be in close proximity to stimulation channels, in most cases, being directly adjacent to the channels. We additionally observed that there was an exponentially decreasing relationship between channel distance from the stimulation electrodes and channel separability and sensitivity performance. These results are consistent with known properties of spatial charge spread following an exponential decrease in power. However, this relationship does not include potential underlying functional connectivity relationships, where functionally connected but distal channels from stimulation electrodes could have been activated. Future work could involve either examining known connectivity relationships related to stimulation regions or looking at potential outliers in the linear fit relationship to start to investigate the contributions of existing connectivity to stimulation response.

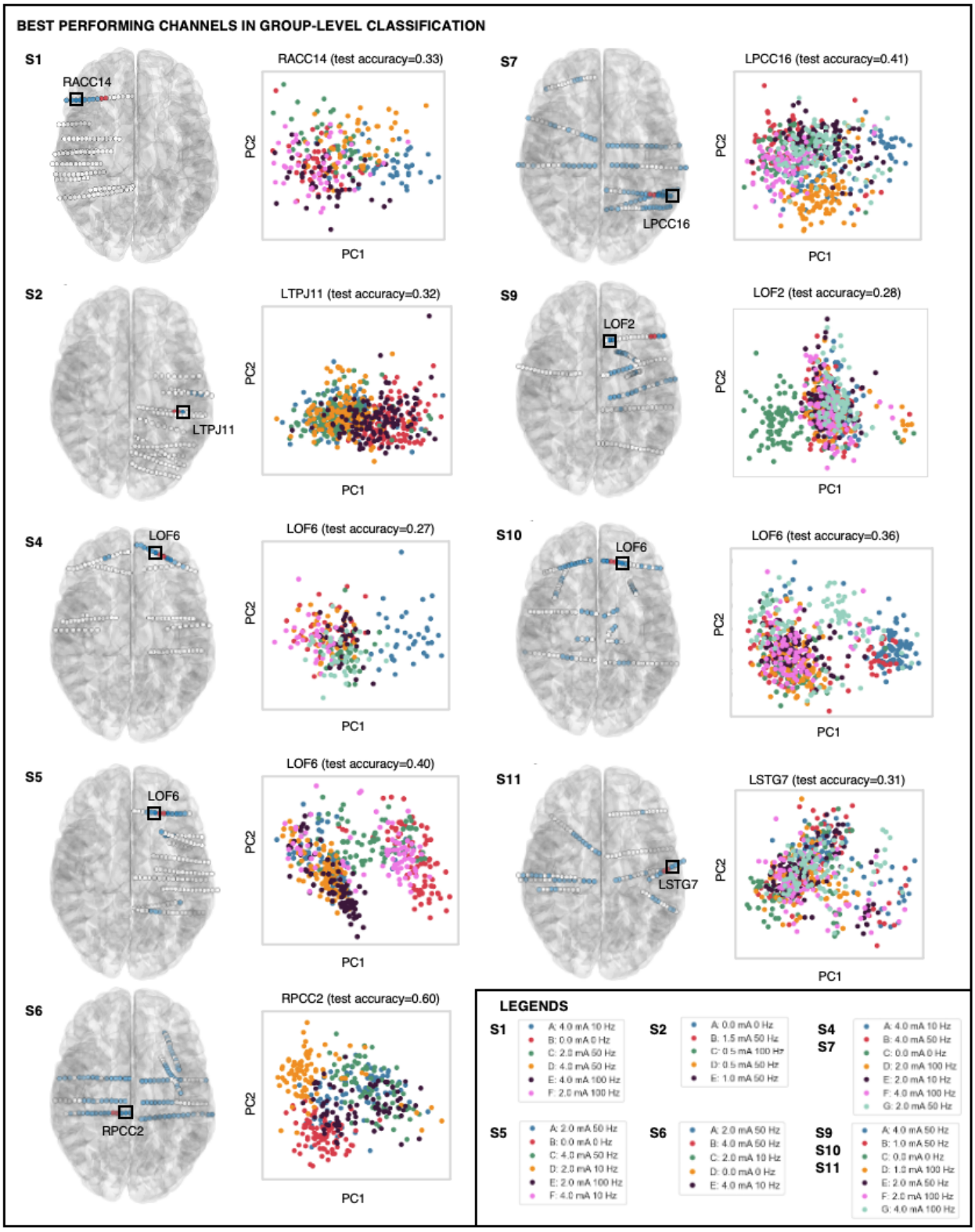
For the duration of our experiment, stimulation responses were variable but remained stable, where there was no observation of drift or compounding effects from continued stimulation. Since the duration of our experiment did not exceed 10 minutes, it's likely that there were no significant state changes occurring in the subject that could contribute to response stability. It's been documented in prior research that response to single pulse stimulation in addition to effectiveness of stimulation intervention varies based on brain state[18,26,42,99]. Additional work is required to document how these stimulation responses differ based on brain state or other contextual differences, such as order of stimulation presentation or potentially shortening the stimulation windows.

Identifying and measuring stimulation separability is an important step towards building adaptive stimulation tools. Achieving separability means that there exist stimulation patterns that can modulate neural activity in potentially different directions, which allows for greater control of modulating specific activity. While we made initial steps of identifying relationships between stimulation parameters and separability, additional work is needed to identify a relationship between stimulation parameters and response to better understand how changing stimulation parameters changes outcomes. Upon building this forward model, one strategy is to reverse the model to determine what stimulation parameters are needed to produce a defined outcome. This modeling strategy has been demonstrated in non-human primates[36], and this chapter lays groundwork for enabling this work in human populations.

6.5 Supplemental Figures

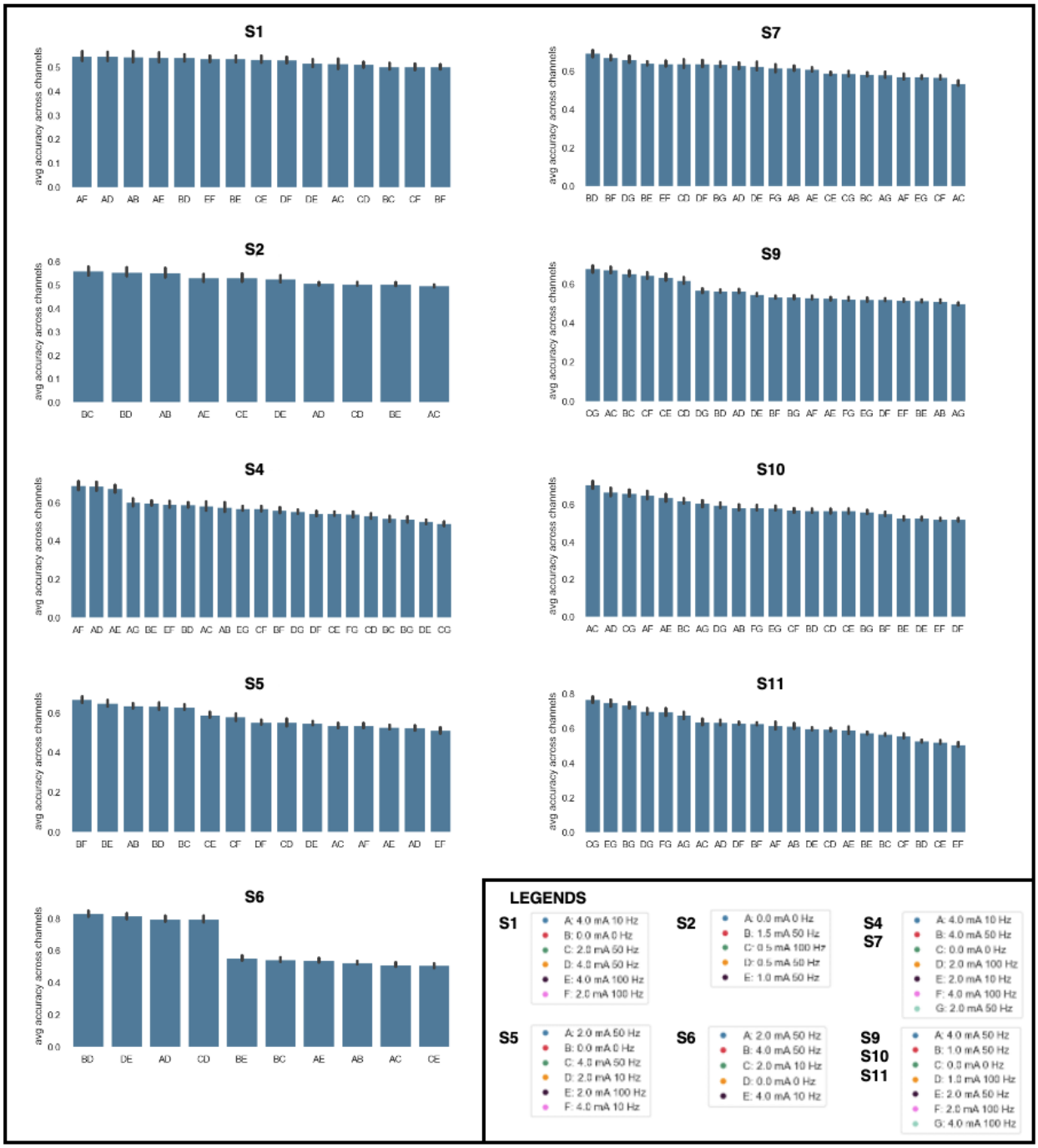


Supplemental Figure 6.1: Electrode montages for all subjects. All subjects possessed more electrodes than could be recorded by our TDT device, so only a subset of electrodes annotated in blue were included in data recordings, and channels left out are annotated in gray. Red electrodes indicate stimulation electrodes for each subject.

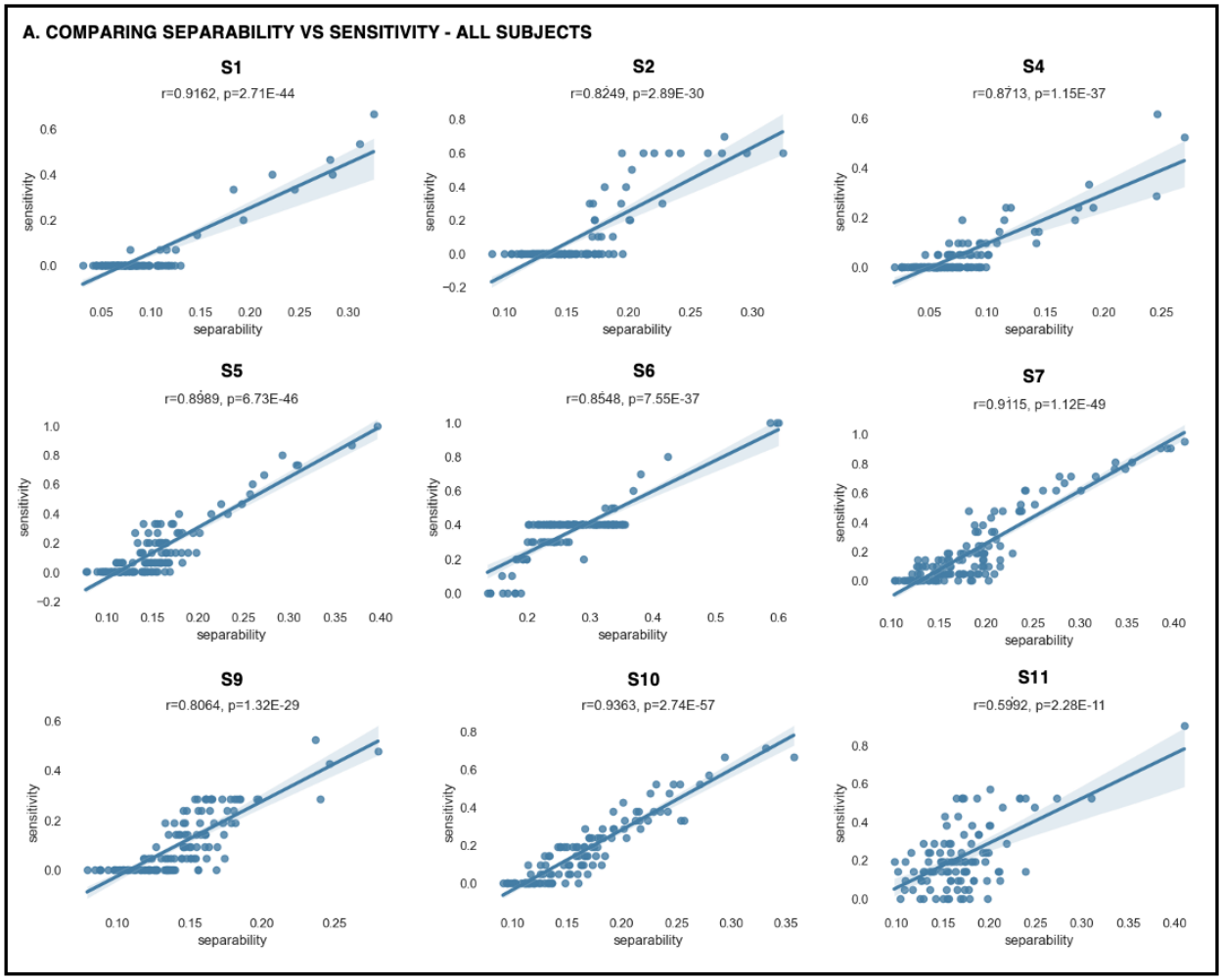


Supplemental Figure 6.2: Best performing channels for each subject in group-level classification task. Best performing channels are highlighted, and its corresponding

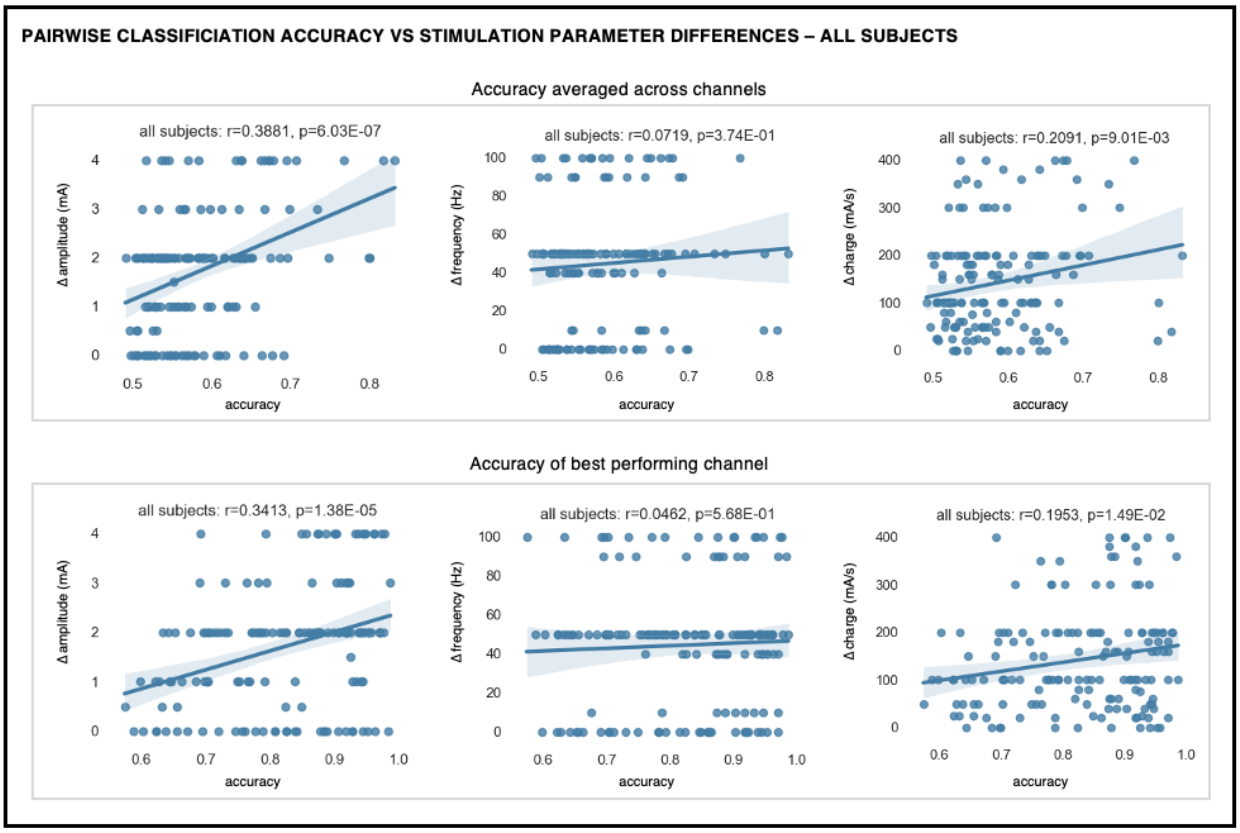
stimulation response in PC space is displayed, reporting test accuracy and color-coded by stimulation condition. Note that stimulation conditions may vary between subjects.



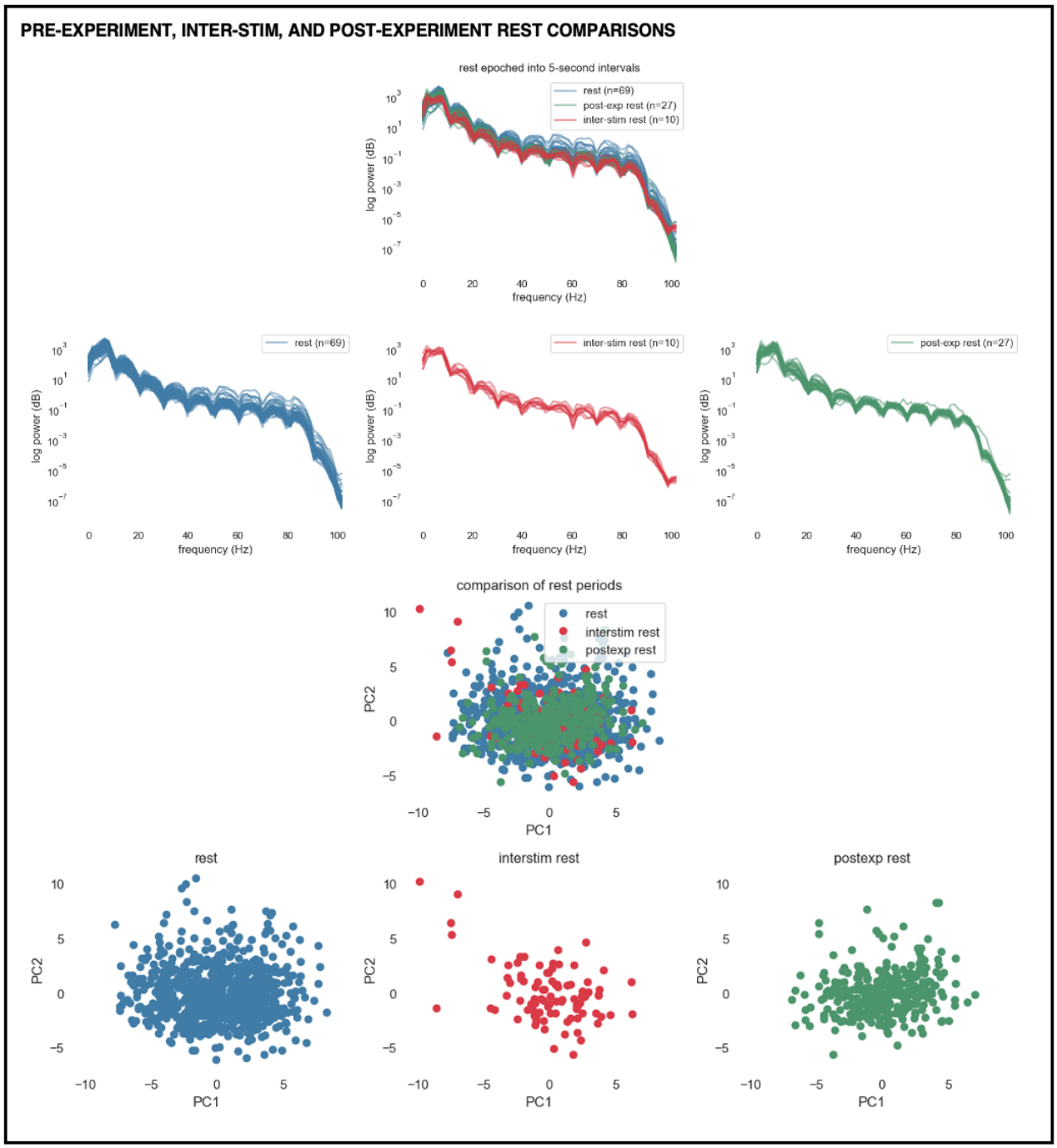
Supplemental Figure 6.3: Stimulation pairs ranked by average separability across channels for each subject.



Supplemental Figure 6.4: High correlation between group-level separability and pairwise sensitivity across subjects, with Pearson correlation coefficient values ranging from 0.8064 to 0.9363 with the exception of subject S11 with a correlation coefficient of 0.5992. Reported p-values test whether the slope is statistically significant from zero, which is the case for all subjects.



Supplemental Figure 6.5: Linear relationship between pairwise classification accuracies and stimulation parameters aggregated across all subjects.



Supplemental Figure 6.6: Pre-experimental rest, inter-stim rest, and post-experimental rest are comparable in frequency behavior in example Subject S7. A non-parametric Mann Whitney U test was performed between rest conditions, and no significant p-values were reported, indicating that these distributions were not statistically significant from each other.

7. Where to stimulate: Mapping regions responsive to neurostimulation

This chapter is based on the following paper in preparation.

[in preparation] Samantha H Sun, Lila H Levinson, Kurt Weaver, Andrew Ko, Ben Grannan, Jason Hauptman, Jeff Tsai, Jeffrey A Herron, Rajesh PN Rao, Jeffrey G Ojemann; Responsiveness to high-frequency electrical stimulation can be predicted using single-pulse stimulation response.

7.1 Introduction

Stimulation therapies rely on selection of optimal stimulation electrodes that produce desirable downstream effects, often improvements in symptom state. Recent work has focused on identifying biomarkers associated with symptom state[9,93,104], which can provide feedback on therapy effectiveness and also be incorporated into closed loop systems to adjust stimulation depending on biomarker state[9]. However, fewer advancements have been made in identifying which target stimulation regions are most effective in improving symptom state.

Identification of stimulation sites for deep brain stimulation has typically relied on a combination of known anatomy and circuitry, lesion studies[46], functional neuroimaging, tractography[47,48], and opportunistic discoveries in related pathologies[49,50]. Despite best efforts in identifying candidate stimulation sites, there remains patient heterogeneity in response to stimulation therapies, particularly in mood disorders[96,105,106]. Patient-specific adjustment of stimulation site and stimulation parameters may be needed to provide sufficient therapy across patients.

In addition, while DBS for tremor suppression tends to provide immediate feedback in terms of stimulation efficacy, DBS for mood disorders may not provide that same feedback and may require the use of neural biomarkers or other indicators to determine stimulation efficacy. With current DBS applications and as new potential applications emerge, methods to identify candidate electrodes for stimulation are needed.

In the previous chapter, we demonstrated a method to identify intracranial electrodes that are sensitive to changes in stimulation input at a particular site. In most cases, only a select few electrodes per subject were sensitive to changes in stimulation. Being able to quickly identify these stimulation-responsive channels will help expedite the decision on whether the stimulation site is effective at modulating symptom-related biomarkers.

In this chapter, we provide evidence that single-pulse stimulation response can predict electrode responsiveness to continuous neurostimulation. We also demonstrate an experimental design to create maps of stimulation response, which can subsequently be used to select stimulation sites based on target regions of interest.

7.2 Methods

Subjects

Subjects and subject demographics were detailed in Chapter 6. Subject S11 was removed due to not having a sufficient number of stimulation trials required to reliably measure an evoked response[107].

Data Acquisition

Data acquisition was performed as outlined in Chapter 2: General methods.

Stimulation protocol

Stimulation waveform design and electrode selection were performed as outlined in Chapter 2: General methods. Stimulation experiments were detailed in Chapter 6.

Single pulse stimulation

We performed repeated, single-pulse stimulation (amplitude=6 mA, frequency=0.5-1 Hz, variable jitter) at the selected stimulation site to obtain a measurement of evoked effective connectivity[26]. Individual subject stimulation channels and stimulation details are listed in Table 7.1. Jitter was used in some experiments to prevent potential entrainment effects.

Table 7.1: Subject-specific stimulation details for evoked potential experiment.

	active recording electrodes	stimulation channels	pulse count	Inter-pulse interval	Jitter %
S1	109	RACC 7-8	200	2000 ms	0
S2	113	LTPJ 9-10	100	1750 ms	15
S4	113	LOF 7-8	200	2000 ms	0
S5	125	LOF 7-8	200	2000 ms	0
S6	125	RPCC 4-5	500	2000 ms	15
S7	126	LPCC 11-12	500	2000 ms	15
S9	124	LOF 13-14	50	2000 ms	15
S10	124	LOF 3-4	50	1000 ms	0

Data processing

Preprocessing methods were performed as outlined in Chapter 2.

Evoked potential amplitude

Evoked potentials (EPs) were averaged across repetitions, and a peak-to-peak amplitude measurement was made using the most positive and most negative point between a window defined as 10 - 80 ms after stimulation (Figure 7.1). Responses were visually inspected to ensure the window effectively captured the early components of the evoked response. After amplitudes were calculated across all channels, the values were min-max normalized such that they were between 0 and 1. Normalized amplitudes were log transformed as doing so helped the dataset have a more normal distribution.

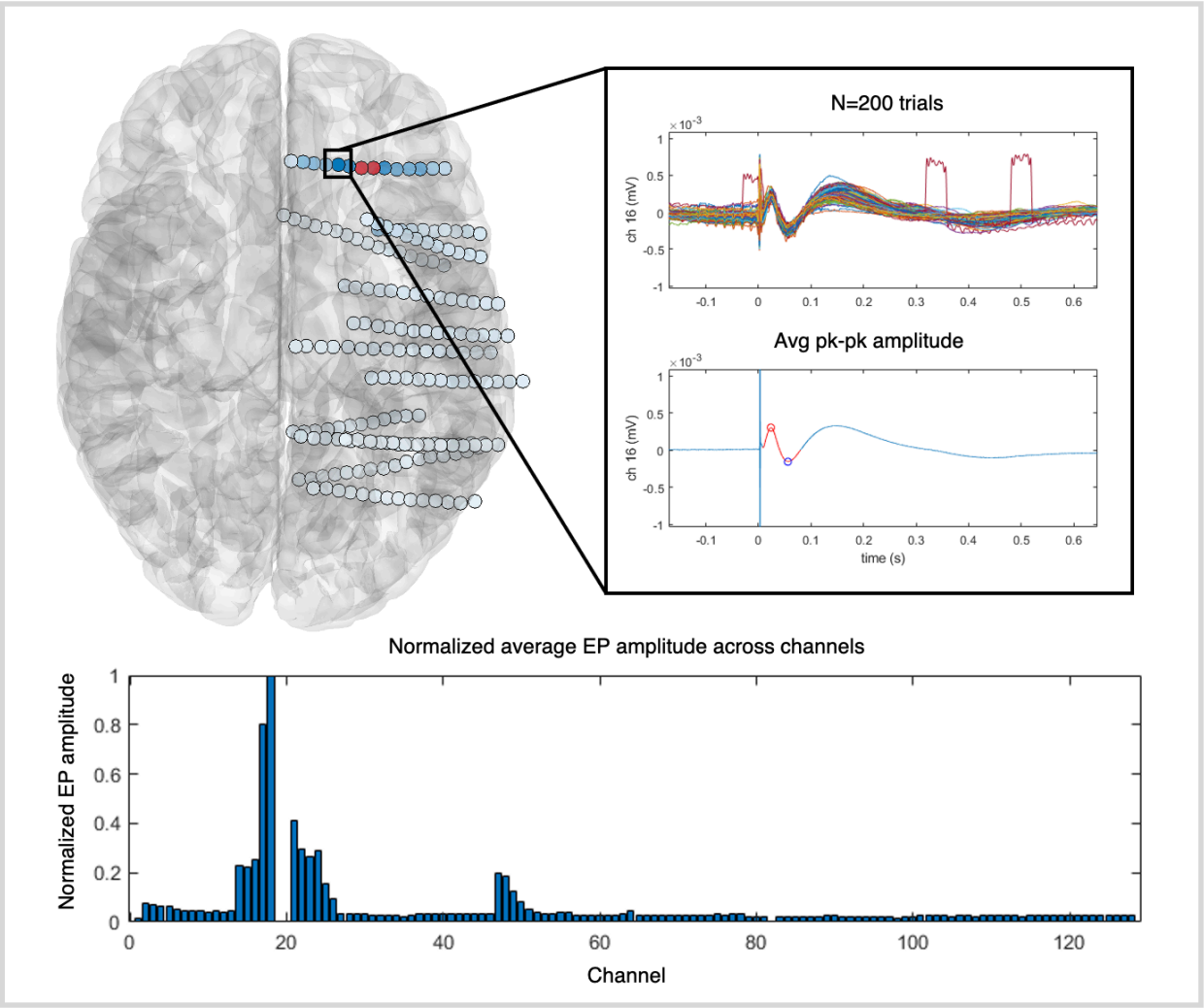


Figure 7.1: Evoked potential amplitudes averaged across repetitions in example subject S5. Stimulation electrodes are shown in red, and a demonstration of calculating the average evoked potential response in a nearby channel is shown on the top right. This process was repeated across all channels, which allowed identification of sites that demonstrated effective connectivity with the stimulation site.

Separability measurements

Separability measurements were calculated with the same methods in Chapter 6.

Distance measurements

For each channel, its Euclidean distance to the point between the stimulation channels was calculated. Due to known properties of charge spread in electrical stimulation, where power

is inversely related to the squared distance[12], we additionally calculated the square root of the Euclidean distance.

Coherence measurements

To establish a benchmark for using EP amplitude as a predictor of channel sensitivity, we included a resting state connectivity metric to our analysis. This metric allowed us to examine potential differences between resting state connectivity and effective connectivity and motivate the use of EP amplitude over non-stimulation predictors.

Resting-state data were collected prior to stimulation experiments. Coherence at canonical frequency bands (delta [0.1-3 Hz], theta [4-7 Hz], alpha [8-12 Hz], beta [13-30 Hz], low gamma [30-55 Hz], high gamma [65-100 Hz]) was calculated for between individual recording channels and the stimulation channel pair using Welch's method.

Linear modeling and prediction

Each predictor, EP amplitude, distance, and coherence, was fit to a linear model using separability as the dependent variable. To assess the model fit, we calculated Pearson correlation coefficient (r) and a p-value to test if the model slope was significantly different than zero, indicating no correlation. Combinations of predictors were also fit for comparison. In addition, a null model using EP amplitude as a predictor but where training labels were scrambled was used as a benchmark.

To compare the prediction performance of these models, we compared mean squared error across our models. We performed an 80/20 train/test split on the dataset, where the linear model was fit on the training dataset only and evaluation was performed on the test data only. This process was repeated 100 times, where the training and test set were randomly reassigned, and the outcome was a distribution of MSE values for each model. A student's t-test was performed to evaluate significance in MSE distributions across models.

7.3 Results

EP amplitude and distance from stimulation are strongly correlated to separability

We observed strong linear relationships between EP amplitude, distance from stimulation channels, and channel separability (Figure 7.2). EP amplitude had a strong positive correlation with separability, and electrode distance had a strong negative correlation with separability. EP amplitude and electrode distance had a strong negative relationship, where electrodes farther from the stimulation pair tended to have smaller EP amplitudes. This observation is in alignment with prior work[18], but it does ignore any potential long-distance synaptic connections between the stimulation site and other regions.

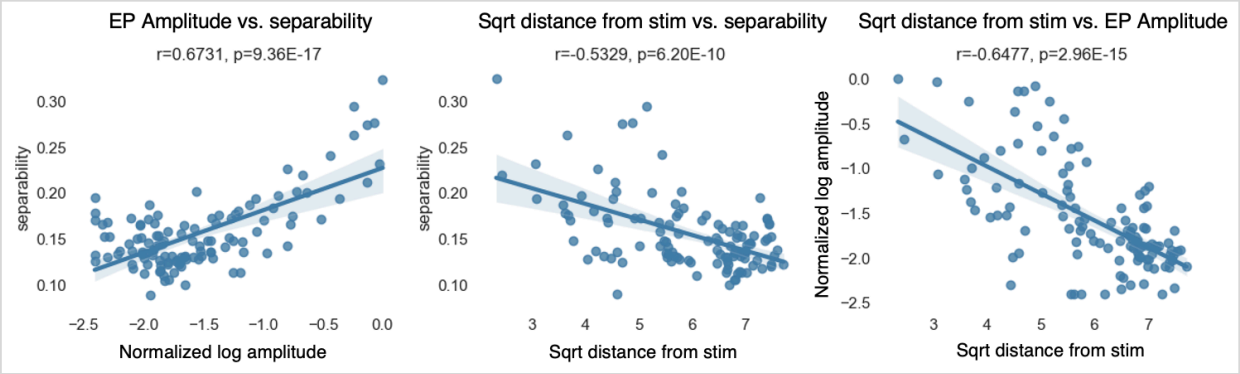


Figure 7.2: Linear relationships between EP amplitude, distance from stimulation channels, and channel separability in example subject S2.

EP amplitude predicts stimulation separability in individual electrodes

While both distance and EP amplitude had strong correlations with separability, in 5 of 8 subjects, EP amplitude outperformed our distance metric in predicting channel separability (Figure 7.3). This important finding indicates that distance alone is not sufficient in modeling a channel's responsiveness to stimulation, and it's likely that EP amplitude contains additional information about the effective, stimulation-relevant connectivity that gives it its improved model performance.

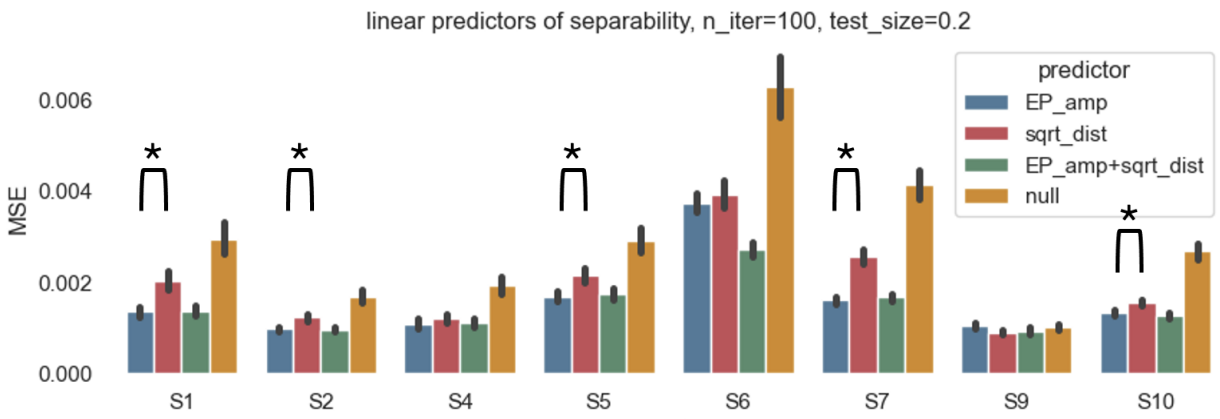


Figure 7.3: Comparison of linear model predictors of stimulation separability across subjects. In five of eight subjects, the EP amplitude model yielded significantly lower errors compared to using distance as a predictor alone. Models all performed better than a null model, except in subject S9.

Combining EP amplitude and distance features yielded similar MSE values to EP amplitude alone in all subjects except in subject S6, where the combined metrics produced significantly lowered MSE values (p-value < 0.05). With this observation, we would recommend continued use of this combined model approach.

EP amplitude outperforms resting state coherence

Coherence measurements perform better than chance in all frequency bands except in the high-gamma band but are outperformed by EP amplitude in predicting separability (Figure 7.4). This result motivates that resting state connectivity alone is also not sufficient in predicting stimulation separability. EP amplitude additionally outperformed combined linear models with coherence and distance metrics, further providing evidence that the most effective predictor of channel separability relies on effective connectivity information.

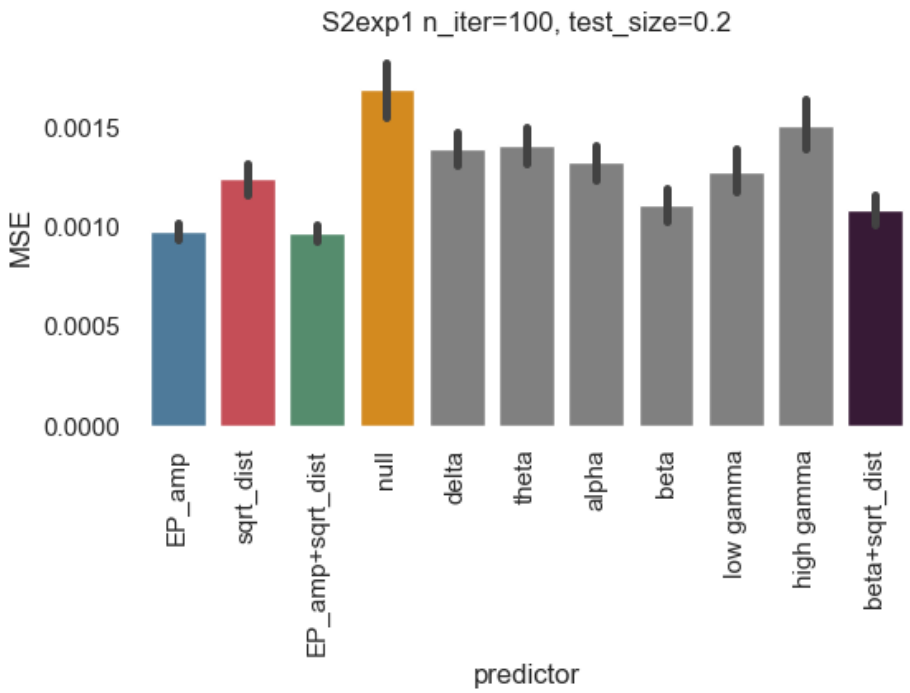


Figure 7.4: Coherence predictors example in subject S2. Coherence in most frequency bands produced an MSE significantly lower than chance, although none had an MSE lower than the EP amplitude model. An example of a combined model, using beta coherence and our distance metric, is still outperformed by EP amplitude.

Mapping stimulation responsive electrodes: an experimental design

As we demonstrated that EP amplitudes were the best predictor of electrodes that were sensitive to changes in stimulation at a given stimulation site, we designed an experimental protocol to build stimulation responsive electrode maps across many stimulation sites. These resulting stimulation maps seek to answer and provide visualizations for the following two questions:

1. Divergent map: Given a stimulation site, what electrodes are predicted to be sensitive to changes in stimulation?
2. Convergent map: Given a target region, what stimulation sites are candidates for modulating this region's behavior?

The first question was already addressed in the prior results, where EP amplitude measurements could be used to identify stimulation responsive electrodes. The second

question may be more relevant for applications in neuromodulation, where a target biomarker has already been identified, and an ideal stimulation site has yet to be selected.

The suggested protocol is as follows:

Single-pulse stimulation experiment: perform single-pulse stimulation at multiple stimulation sites at clinically safe stimulation amplitudes, at minimum 10 pulses[107]-per stimulation site.

Calculating EP amplitudes: For each stimulation site, measure the peak-to-peak evoked potential response and calculate the average amplitude response across all electrodes.

Threshold for large EP responses: Normalize EP amplitude responses, convert to a logarithmic scale, and set an appropriate threshold cutoff for identifying significant EP amplitudes (we used a cutoff of 2 standard deviations larger than the mean).

Normalization recommendations: We express caution that proper normalization is critical in order to obtain correct results. For the divergent map, we recommend performing EP amplitude normalization across recording channels to identify channels that produced the set of largest EP responses. In contrast, for the convergent map, we recommend first grouping EP amplitude responses by stimulation site before performing normalization. This initial grouping ensures that EP amplitudes are properly ranked.

Upon curating this dataset, we created visualizations that indicated responsive electrodes given a stimulation site, and candidate stimulation sites given a target region of interest (Figure 7.5).

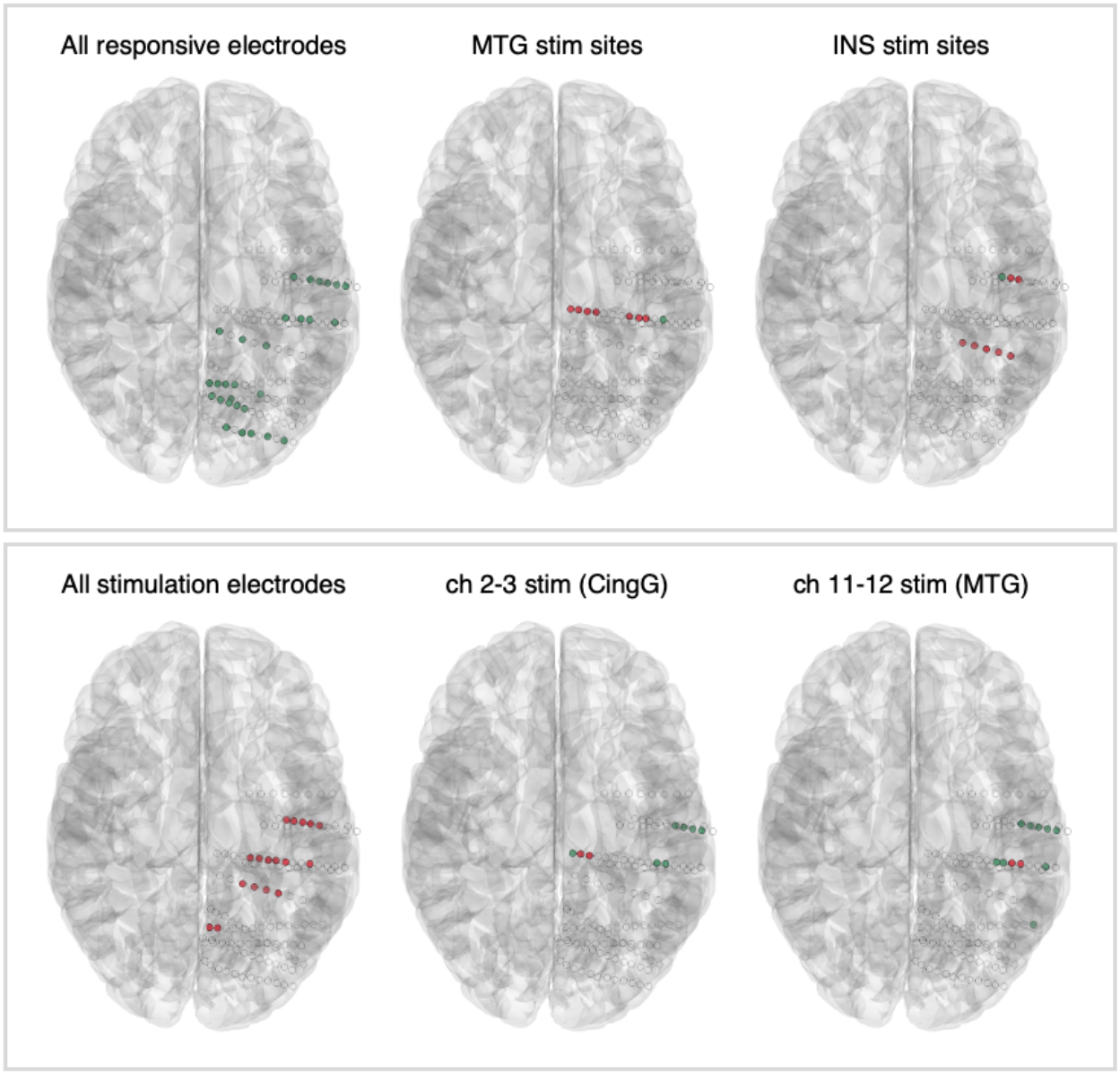


Figure 7.5: Convergent (top) and divergent (bottom) map examples in subject S2, where stimulation sites are in red and responsive sites are in green. All responsive electrodes (top left) were sites that produced large EP amplitude responses to any of the stimulation sites (bottom left). Example target regions and their respective stimulation candidate sites are displayed (top middle, right), and example stimulation sites and their respective responsive electrodes are displayed (bottom middle, right). (MTG = middle temporal gyrus, INS = insular cortex, CingG = cingulate gyrus)

7.4 Discussion

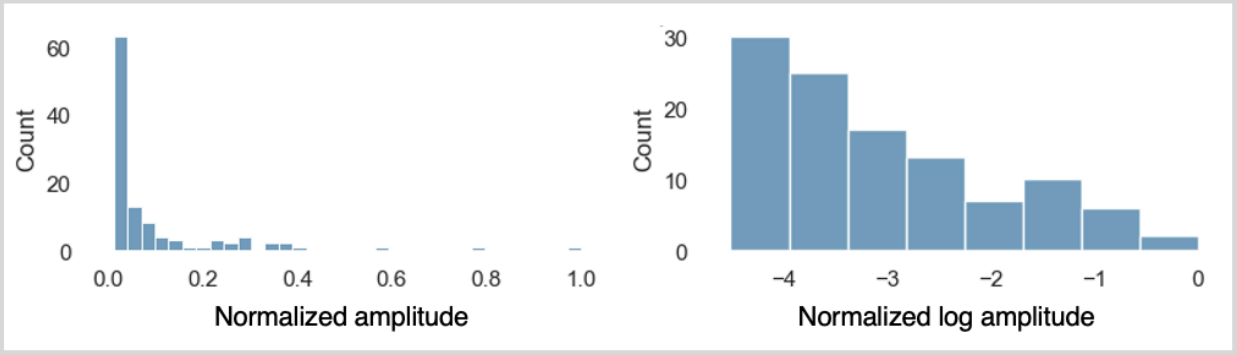
In this chapter, we demonstrated that evoked potential amplitude was the best predictor of an electrode's sensitivity to changing stimulation parameters, where a linear model using only the evoked potential amplitude as a feature outperformed models utilizing resting state connectivity and electrode distance metrics. We additionally identified that linear models using a combination of evoked potential amplitude and distance as features would provide the best performance across all subjects. Leveraging these results, we proposed an experimental paradigm to build stimulation response maps across many stimulation sites and demonstrated example visualizations of how these stimulation maps could be used in the context of identifying stimulation candidates for targeted neuromodulation.

In contrast to resting state connectivity, effective connectivity is a connectivity measure specifically in the context of stimulation, and it has been used to predict outcomes in stimulation entrainment experiments[27]. While single-pulse stimulation mapping is already a known method for studying effective connectivity across brain regions[27,108,109], this work confirms and extends prior findings, where effective connectivity can additionally predict region sensitivity to different stimulation inputs. This contribution provides evidence that effective connectivity mapping can be used to inform stimulation sites effective for neuromodulation.

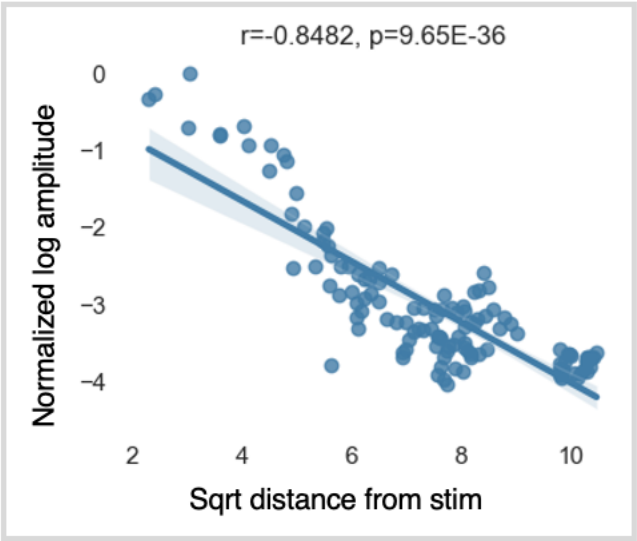
Our intended use case for the stimulation mapping experiment is to inform stimulation site selection for neuromodulation therapies, including in deep brain stimulation. Due to the heterogeneity of patient response to neuromodulation therapies, especially in treatment-resistant mood disorders[96,105,106], patient-specific stimulation protocols and regions are likely needed. Adaptive stimulation therapies, where stimulation is adjusted based on ongoing neurofeedback, has demonstrated initial success as a patient-specific stimulation protocol[110]. This chapter contributes to the process of selecting patient-specific stimulation electrodes, specifically ones that can produce variable responses in target regions of interest. Selected candidate electrodes will need to be tested to determine if stimulation produced therapeutic outcomes, which could be measured using direct behavioral reports or by monitoring symptom-related biomarkers.

This proposed stimulation protocol will provide the most informative maps in cases where the patient has multiple electrode implants, such as in intractable epilepsy patients undergoing monitoring with temporary electrode implants of upwards of 200 total electrode contacts. Receiving temporary implants for monitoring or screening purposes has been growing in usage for clinical trials for treatment-resistant depression[9,111], in addition to DBS implant surgeries, where a temporary strip of surface electrodes is placed on the cortex for research purposes[112,113]. As these practices become more common, we can envision a standard where stimulation mapping across multiple stimulation sites is performed, in addition to common practices of recording baseline resting state and single-site EP recordings[114].

7.5 Supplemental Figures



Supplemental Figure 7.1: Log transformation linearized distribution of EP amplitudes. Many channels had small EPs, and few channels had large EPs that were magnitudes greater than the small EPs. A log transformation helped linearize the distribution of EPs for improved regression analysis.



Supplemental Figure 7.2: Log EP amplitude increased linearly with square-root of electrode distance.

8. Conclusion

In this thesis, we contributed to three components of closed-loop neuromodulation therapies: brain-state estimation, characterizing patterned stimulation response, and mapping regions sensitive to stimulation patterns.

In Chapters 4 and 5, using neural spectral features collected from human epilepsy patients, we built two independent brain state estimation models for sleep/wake classification and pain state classification. In Chapter 4, we trained an unsupervised machine learning model that had specific synergy with sleep/wake classification and demonstrated that it outperformed other unsupervised methods. We also identified subject-specific biomarkers of pain in Chapter 5 and used them to make predictions of high vs low pain state and pain on a 0-10 scale. In the context of adaptive therapies, these models can be used for real-time brain state estimation to help determine whether stimulation therapy is needed. The pain state model can indicate when a patient's pain levels exceed a threshold, which would inform closed-loop neuromodulation systems to start administering stimulation therapy. While the sleep/wake model does not predict a specific disorder or disease state, it may still provide critical information in adaptive therapies, where therapy may or may not be needed during sleep depending on the disorder.

In Chapter 6, we additionally developed a framework for characterizing stimulation response with patterned stimulation, identifying regions that are sensitive to changes in stimulation and which stimulation patterns are more easily separable. This work contributes to our understanding of neuromodulation using electrical stimulation, where we identified that stimulation amplitude seemed to be the only stimulation parameter that could impact response. We also identified that there tended to be few sites that were sensitive to stimulation across subjects, which means for any given stimulation site, there may only be a few targets capable of being modulated. Towards building adaptive closed-loop neuromodulation therapies, this work confirms that patterned stimulation can produce distinct outcomes, and additional work is needed in building an explicit model between stimulation parameters and response, which may include increasing the number of parameters to be explored.

Finally, in Chapter 7 we showed that single-pulse stimulation can be used as an efficient way to identify electrodes that are sensitive to patterned stimulation and demonstrated a proof-of-concept experiment design which can generate a stimulation response map. These maps can then allow researchers to quickly identify candidate stimulation electrodes to modulate the activity of target regions. Stimulation response mapping has already been well established as a method for measuring connectivity between regions[27], and our work takes it one step farther by creating a lookup database that produces stimulation electrode suggestions for all potential neuromodulation targets. In addition, while previous work identified that single-pulse stimulation response can be used to predict stimulation intervention outcomes, our work adds to the utility of single-pulse stimulation, where it can also be used to identify regions sensitive to changes in stimulation.

8.1 Towards adaptive closed loop stimulation

From stimulation parameters to stimulation response

One open area of research involves building an explicit model between stimulation input and stimulation response. There are many factors that increase the complexity or difficulty of this problem, including the brain being a non-stationary system and the stimulation parameter space being vast and multidimensional. Recent efforts in building this relationship typically will constrain stimulation parameters to one to two dimensions[34,36], and future work is needed to more efficiently explore this parameter space. State-space modeling of intrinsic brain activity has been shown to be successful at predicting mood states in humans[104] and predicting stimulation outcomes in non-human primates[36], and we believe that these models can be effective in human stimulation models as well, given sufficient data. Other promising methods include using reinforcement learning or control theory models to determine which stimulation inputs can drive neural activity towards a desired state, allowing the development of "neural co-processors"[115–117]. As opportunities for human intracranial research are limited, models that are capable of learning with potentially limited datasets or in short time-frames may be preferred.

8.2 Challenges and limitations

Work performed in an epilepsy patient population

All our findings involved research performed in epilepsy patients, and while we avoided stimulating and recording at seizure onset regions, we assumed that activity outside of seizure events were comparable to that of a healthy brain. Further work is needed to confirm the generalizability of these findings to healthy brains and in patients with other neurological conditions intended for aDBS therapies. As electrode locations were determined by clinical needs, we did not have consistent electrode coverage across patients which made group-level comparisons challenging.

Translating biomarkers to behavioral outcomes

Our efforts in identifying biomarkers that could information stimulation therapies contribute to many other ongoing efforts in biomarker detection. While biomarkers may be indicators for symptom state, therapies that seek to change biomarker activity may not necessarily result in changes in behavioral outcome. Another possibility is that stimulation could potentially change behavioral outcomes in undesirable ways[7]. One strategy is to study the link between biomarker activity and disease circuitry, and we hypothesize that biomarker activity that is directly involved in disease circuitry will have better stimulation outcomes compared to activity that is produced as a downstream effect. Ultimately, comprehensive testing is needed to determine whether a biomarker is an effective indicator for stimulation-induced behavioral outcomes. This testing process could be simple in cases where behavior closely follows stimulation, such as in tremor, but may be more challenging in cases where stimulation therapy requires longer durations of time before observing meaningful behavioral outcomes.

Stimulation artifact contamination

Removal of stimulation artifact is high priority in stimulation work, as remaining artifact could produce false results, especially in the case of separability of stimulation response. While great efforts were taken to remove stimulation artifact, large artifact presence remained in the frequency space and constrained the range of frequency bands we could use. Frequency bins surrounding stimulation frequencies were removed, and in the case where 10 Hz stimulation was used, a large proportion of frequency bins were removed from analysis,

such that key features may have been obscured by stimulation artifact. While strategies exist to avoid analyzing active stimulation periods, such as introducing a rest period between stimulation presentations, uncovering ongoing dynamics during stimulation could introduce great strides in our basic understanding of the mechanisms of stimulation to better wield stimulation as a tool for neuromodulation.

8.3 Neuroethics

While there are many benefits arising from moving towards an adaptive, closed-loop, personalized medicine therapy space, there are similarly an increasing number of ethical considerations that must be carefully handled as we move forward in this space.

Adaptive therapies could offer automatic therapy without patient needing to turn the device on or off, which provides a greater level of convenience for the patient. However, this could also affect the patient's sense of agency and introduce a feeling of having lack of control[118-120]. Furthermore, when interviewing DBS patients being treated for either major depressive disorder or obsessive-compulsive disorder, some expressed concern about how their device may alter their sense of self, either by producing artificial emotions or being unable to tell whether the device was responsible for their current emotions[121,122]. Continued efforts must be made in collaboration with patients receiving adaptive therapies to help define what a clinically successful treatment may be and how agency fits within that definition.

Deciding between open-loop, patient-controlled, or closed-loop stimulation strategies can impact both clinical outcome and patient experience. When asked about having the ability to control their own DBS treatment, most patients preferred to have minimal control of their device[121], with the primary concern being misuse or potential addiction to the treatment. Additionally, adaptive therapies are not guaranteed to work perfectly, and potential ethical implications could arise when patients are either undertreated or overtreated from adaptive therapies. For instance, DBS patients report that family members were quick to assign blame to the device in the event of interpersonal disagreements, when it was unknown whether the device was undertreating or if it was an authentic interaction[121]. This assignment of responsibility and blame could also convolute legal cases, where it becomes unclear whether a patient's actions can be entirely assigned as their own.

With an increasing application space, we also should consider the risk of introducing bias in developing neurotechnologies, where devices and/or adaptive algorithms may be more effective in certain populations over others. One example of how bias is present in this space is when considering the socioeconomic cost of receiving DBS— specifically how it impacts the demographics of individuals participating in research and whether this population can be representative of people from all socioeconomic backgrounds. There is a complex intersection between socioeconomic status, demographics, and underserved populations, and the interplay of these factors may influence the quality of care as these technologies become accessible to these populations. This area of research is understudied and provides nuance to our findings and findings of many research areas pertaining to DBS and intracranial studies.

9. References

1. Groiss, S. J., Wojtecki, L., Südmeyer, M. & Schnitzler, A. Review: Deep brain stimulation in Parkinson's disease. *Ther Adv Neurol Disord* **2**, 379–391 (2009).
2. Jarosiewicz, B. & Morrell, M. The RNS System: brain-responsive neurostimulation for the treatment of epilepsy. *Expert Review of Medical Devices* **18**, 129–138 (2021).
3. Rowald, A. *et al*. Activity-dependent spinal cord neuromodulation rapidly restores trunk and leg motor functions after complete paralysis. *Nat Med* **28**, 260–271 (2022).
4. Herron, J. A. *et al*. Cortical Brain–Computer Interface for Closed-Loop Deep Brain Stimulation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25**, 2180–2187 (2017).
5. Ferleger, B. I. *et al*. Fully implanted adaptive deep brain stimulation in freely moving essential tremor patients. *J. Neural Eng.* **17**, 056026 (2020).
6. Shanechi, M. M. Brain–machine interfaces from motor to mood. *Nat Neurosci* **22**, 1554–1564 (2019).
7. Scangos, K. W., Makhoul, G. S., Sugrue, L. P., Chang, E. F. & Krystal, A. D. State-dependent responses to intracranial brain stimulation in a patient with depression. *Nat Med* **27**, 229–231 (2021).
8. Oehrn, C. R. *et al*. Personalized chronic adaptive deep brain stimulation outperforms conventional stimulation in Parkinson's disease. *medRxiv* 2023.08.03.23293450 (2023) doi:10.1101/2023.08.03.23293450.
9. Scangos, K. W. *et al*. Closed-loop neuromodulation in an individual with treatment-resistant depression. *Nat Med* **27**, 1696–1700 (2021).
10. Grover, S., Nguyen, J. A., Viswanathan, V. & Reinhart, R. M. G. High-frequency neuromodulation improves obsessive–compulsive behavior. *Nat Med* **27**, 232–238 (2021).

11. Orsborn, A. L. *et al*. Closed-Loop Decoder Adaptation Shapes Neural Plasticity for Skillful Neuroprosthetic Control. *Neuron* **82**, 1380–1393 (2014).

12. Caldwell, D. J., Ojemann, J. G. & Rao, R. P. N. Direct Electrical Stimulation in Electrocorticographic Brain–Computer Interfaces: Enabling Technologies for Input to Cortex. *Frontiers in Neuroscience* **13**, (2019).

13. Papasavvas, C. A. *et al*. Band power modulation through intracranial EEG stimulation and its cross-session consistency. *J. Neural Eng.* **17**, 054001 (2020).

14. Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* **117**, 500–544 (1952).

15. Burkitt, A. N. A Review of the Integrate-and-fire Neuron Model: I. Homogeneous Synaptic Input. *Biol Cybern* **95**, 1–19 (2006).

16. Caporale, N. & Dan, Y. Spike timing-dependent plasticity: a Hebbian learning rule. *Annu Rev Neurosci* **31**, 25–46 (2008).

17. Shupe, L. & Fetz, E. An Integrate-and-Fire Spiking Neural Network Model Simulating Artificially Induced Cortical Plasticity. *eNeuro* **8**, ENEURO.0333-20.2021 (2021).

18. Keller, C. J. *et al*. Intrinsic functional architecture predicts electrically evoked responses in the human brain. *Proceedings of the National Academy of Sciences* **108**, 10308–10313 (2011).

19. Solomon, E. A. *et al*. Medial temporal lobe functional connectivity predicts stimulation-induced theta power. *Nat Commun* **9**, 4437 (2018).

20. Fox, K. C. R. *et al*. Intrinsic network architecture predicts the effects elicited by intracranial electrical stimulation of the human brain. *Nat Hum Behav* **4**, 1039–1052 (2020).

21. Papasavvas, C., Taylor, P. N. & Wang, Y. *Long-Term Changes in Functional Connectivity Predict Responses to Intracranial Stimulation of the Human Brain*. http://arxiv.org/abs/2105.02805 (2021) doi:10.48550/arXiv.2105.02805.

22. Bloch, J. *et al*. Cortical network structure mediates response to stimulation: an optogenetic study in non-human primates. 2021.05.17.444526 Preprint at https://doi.org/10.1101/2021.05.17.444526 (2021).

23. Khambhati, A. N., Shafi, A., Rao, V. R. & Chang, E. F. Long-term brain network reorganization predicts responsive neurostimulation outcomes for focal epilepsy. *Sci. Transl. Med.* **13**, eabf6588 (2021).

24. Fan, J. M. *et al*. Network connectivity predicts effectiveness of responsive neurostimulation in focal epilepsy. *Brain Communications* fcac104 (2022) doi:10.1093/braincomms/fcac104.

25. Huang, Y. *et al*. Intracortical Dynamics Underlying Repetitive Stimulation Predicts Changes in Network Connectivity. *J. Neurosci.* **39**, 6122–6135 (2019).

26. Keller, C. J. *et al*. Mapping human brain networks with cortico-cortical evoked potentials. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130528 (2014).

27. Keller, C. J. *et al*. Induction and Quantification of Excitability Changes in Human Cortical Networks. *J. Neurosci.* **38**, 5384–5398 (2018).

28. Jin, M., Beck, J. M. & Glickfeld, L. L. Neuronal Adaptation Reveals a Suboptimal Decoding of Orientation Tuned Populations in the Mouse Visual Cortex. *J. Neurosci.* **39**, 3867–3881 (2019).

29. Graczyk, E. L., Delhaye, B. P., Schiefer, M. A., Bensmaia, S. J. & Tyler, D. J. Sensory adaptation to electrical stimulation of the somatosensory nerves. *J. Neural Eng.* **15**, 046002 (2018).

30. Johnson, L. A. *et al*. Direct electrical stimulation of somatosensory cortex in humans using electrocorticography electrodes: a qualitative and quantitative report. *J Neural Eng* **10**, 036021 (2013).

31. Hiremath, S. V. *et al*. Human perception of electrical stimulation on the surface of somatosensory cortex. *PLOS ONE* **12**, e0176020 (2017).

32. Kramer, D. R. *et al*. Functional Frequency Discrimination From Cortical Somatosensory Stimulation in Humans. *Frontiers in Neuroscience* **13**, (2019).

33. Bjånes, D. A. & Moritz, C. T. A Robust Encoding Scheme for Delivering Artificial Sensory Information via Direct Brain Stimulation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27**, 1994–2004 (2019).

34. Bolus, M. F., Willats, A. A., Rozell, C. J. & Stanley, G. B. State-space optimal feedback control of optogenetically driven neural activity. *J. Neural Eng.* **18**, 036006 (2021).

35. Tafazoli, S. *et al*. Learning to Control the Brain through Adaptive Closed-Loop Patterned Stimulation. 2020.03.14.992198 Preprint at https://doi.org/10.1101/2020.03.14.992198 (2020).

36. Yang, Y. *et al*. Modelling and prediction of the dynamic responses of large-scale brain networks during direct electrical stimulation. *Nat Biomed Eng* **5**, 324–345 (2021).

37. Swann, N. C. *et al*. Adaptive deep brain stimulation for Parkinson's disease using motor cortex sensing. *J Neural Eng* **15**, 046006 (2018).

38. Little, S. *et al*. Adaptive deep brain stimulation in advanced Parkinson disease. *Annals of Neurology* **74**, 449–457 (2013).

39. Alagapan, S. *et al*. Cingulate dynamics track depression recovery with deep brain stimulation. *Nature* **622**, 130–138 (2023).

40. Shirvalkar, P. *et al*. First-in-human prediction of chronic pain state using intracranial neural biomarkers. *Nat Neurosci* **26**, 1090–1099 (2023).

41. Usami, K. *et al*. Sleep modulates cortical connectivity and excitability in humans: Direct evidence from neural activity induced by single-pulse electrical stimulation. *Hum Brain Mapp* **36**, 4714–4729 (2015).

42. Arbune, A. A. *et al*. Sleep modulates effective connectivity: A study using intracranial stimulation and recording. *Clin Neurophysiol* **131**, 529–541 (2020).

43. Velisar, A. *et al*. Dual threshold neural closed loop deep brain stimulation in Parkinson disease patients. *Brain Stimul* **12**, 868–876 (2019).

44. Bittar, R. G. *et al*. Deep brain stimulation for pain relief: a meta-analysis. *J Clin Neurosci* **12**, 515–519 (2005).

45. Scangos, K. W., State, M. W., Miller, A. H., Baker, J. T. & Williams, L. M. New and emerging approaches to treat psychiatric disorders. *Nat Med* **29**, 317–333 (2023).

46. Miguel, E. C. *et al*. Evolution of gamma knife capsulotomy for intractable obsessive-compulsive disorder. *Mol Psychiatry* **24**, 218–240 (2019).

47. Widge, A. S. *et al*. Patient-specific connectomic models correlate with, but do not reliably predict, outcomes in deep brain stimulation for obsessive-compulsive disorder. *Neuropsychopharmacology* **47**, 965–972 (2022).

48. Baldermann, J. C. *et al*. Connectivity Profile Predictive of Effective Deep Brain Stimulation in Obsessive-Compulsive Disorder. *Biol Psychiatry* **85**, 735–743 (2019).

49. Goodman, W. K. *et al*. Deep brain stimulation for intractable obsessive compulsive disorder: pilot study using a blinded, staggered-onset design. *Biol Psychiatry* **67**, 535–542 (2010).

50. Luyten, L., Hendrickx, S., Raymaekers, S., Gabriëls, L. & Nuttin, B. Electrical stimulation in the bed nucleus of the stria terminalis alleviates severe obsessive-compulsive disorder. *Mol Psychiatry* **21**, 1272–1280 (2016).

51. Maintz, J. B. A. & Viergever, M. A. A survey of medical image registration. *Medical Image Analysis* **2**, 1–36 (1998).

52. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage* **9**, 179–194 (1999).

53. Kamali, G. *et al*. Transfer Function Models for the Localization of Seizure Onset Zone From Cortico-Cortical Evoked Potentials. *Front. Neurol.* **11**, (2020).

54. David, O. *et al*. Probabilistic functional tractography of the human cortex. *NeuroImage* **80**, 307–317 (2013).

55. Crowther, L. J. *et al*. A quantitative method for evaluating cortical responses to electrical stimulation. *Journal of Neuroscience Methods* **311**, 67–75 (2019).

56. Hashimoto, T., Elder, C. M. & Vitek, J. L. A template subtraction method for stimulus artifact removal in high-frequency deep brain stimulation. *Journal of Neuroscience Methods* **113**, 181–186 (2002).

57. Erez, Y., Tischler, H., Moran, A. & Bar-Gad, I. Generalized framework for stimulus artifact removal. *Journal of Neuroscience Methods* **191**, 45–59 (2010).

58. Mouthaan, B. E. *et al*. Single Pulse Electrical Stimulation to identify epileptogenic cortex: Clinical information obtained from early evoked responses. *Clin Neurophysiol* **127**, 1088–1098 (2016).

59. Sun, L. & Hinrichs, H. Moving average template subtraction to remove stimulation artefacts in EEGs and LFPs recorded during deep brain stimulation. *Journal of Neuroscience Methods* **266**, 126–136 (2016).

60. Trebaul, L. *et al*. Stimulation artifact correction method for estimation of early cortico-cortical evoked potentials. *Journal of Neuroscience Methods* **264**, 94–102 (2016).

61. Caldwell, D. J. *et al*. Signal recovery from stimulation artifacts in intracranial recordings with dictionary learning. *J. Neural Eng.* **17**, 026023 (2020).

62. Palopoli-Trojani, K. *et al.* Temporally non-regular patterns of deep brain stimulation (DBS) enhance assessment of evoked potentials while maintaining motor symptom management in Parkinson's disease (PD). *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation* **16**, 1630–1642 (2023).

63. Jech, R. *et al.* Deep brain stimulation of the subthalamic nucleus affects resting EEG and visual evoked potentials in Parkinson's disease. *Clinical Neurophysiology* **117**, 1017–1028 (2006).

64. Allen, D. P., Stegemöller, E. L., Zadikoff, C., Rosenow, J. M. & MacKinnon, C. D. Suppression of deep brain stimulation artifacts from the electroencephalogram by frequency-domain Hampel filtering. *Clinical Neurophysiology* **121**, 1227–1232 (2010).

65. Lio, G., Thobois, S., Ballanger, B., Lau, B. & Boulinguez, P. Removing deep brain stimulation artifacts from the electroencephalogram: Issues, recommendations and an open-source toolbox. *Clinical Neurophysiology* **129**, 2170–2185 (2018).

66. Chang, J.-Y. *et al.* Multivariate autoregressive models with exogenous inputs for intracerebral responses to direct electrical stimulation of the human brain. *Front. Hum. Neurosci.* **6**, (2012).

67. Welch, P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics* **15**, 70–73 (1967).

68. Rasch, B. & Born, J. About Sleep's Role in Memory. *Physiological Reviews* **93**, 681–766 (2013).

69. Méndez, M. & Radtke, R. A. Interactions Between Sleep and Epilepsy. *Journal of Clinical Neurophysiology* **18**, 106–127 (2001).

70. Dorris, L., Scott, N., Zuberi, S., Gibson, N. & Espie, C. Sleep problems in children with neurological disorders. *Developmental Neurorehabilitation* **11**, 95–114 (2008).

71. Roth, T. *et al*. Sleep Problems, Comorbid Mental Disorders, and Role Functioning in the National Comorbidity Survey Replication. *Biological Psychiatry* **60**, 1364–1371 (2006).

72. Berry, R. B. *et al*. Rules for Scoring Respiratory Events in Sleep: Update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. *Journal of Clinical Sleep Medicine* **08**, 597–619 (2012).

73. Danker-hopfe, H. *et al*. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. *Journal of Sleep Research* **18**, 74–84 (2009).

74. Sun, H. *et al*. Large-Scale Automated Sleep Staging. *Sleep* **40**, (2017).

75. Längkvist, M., Karlsson, L. & Loutfi, A. Sleep Stage Classification Using Unsupervised Feature Learning. *Advances in Artificial Neural Systems* **2012**, 1–9 (2012).

76. Kremen, V. *et al*. Automated unsupervised behavioral state classification using intracranial electrophysiology. *Journal of Neural Engineering* **16**, 026004 (2019).

77. Johnson, M. J. & Willsky, A. S. Bayesian Nonparametric Hidden Semi-Markov Models. (2012).

78. Buzsaki, G. *Rhythms of the Brain*. (Oxford University Press, 2006).

79. Yu, S.-Z. Hidden semi-Markov models. *Artificial Intelligence* **174**, 215–243 (2010).

80. Manning, C. D. *Introduction to Information Retrieval*. (Cambridge University Press, 2008).

81. Forney, G. D. The viterbi algorithm. *Proceedings of the IEEE* **61**, 268–278 (1973).

82. Walsh, B. Markov Chain Monte Carlo and Gibbs Sampling. (2004).

83. Wang, N. X. R., Olson, J. D., Ojemann, J. G., Rao, R. P. N. & Brunton, B. W. Unsupervised Decoding of Long-Term, Naturalistic Human Neural Recordings with Automated Video and Audio Annotations. *Front. Hum. Neurosci.* **10**, (2016).

84. Morton, D. L., Sandhu, J. S. & Jones, A. K. Brain imaging of pain: state of the art. *J Pain Res* **9**, 613–624 (2016).

85. Shirvalkar, P. *et al*. A deep brain stimulation trial period for treating chronic pain. *Journal of clinical medicine* **9**, 3155 (2020).

86. Pham, T. V. Classification of Pain Intensity using Functional Connectivity Networks Derived from Intracranial Electroencephalography in Humans. (2023).

87. Raja, S. N. *et al*. The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. *PAIN* **161**, 1976 (2020).

88. Vowles, K. E. *et al*. Rates of opioid misuse, abuse, and addiction in chronic pain: a systematic review and data synthesis. *Pain* **156**, 569–576 (2015).

89. Shirvalkar, P., Veuthey, T. L., Dawes, H. E. & Chang, E. F. Closed-Loop Deep Brain Stimulation for Refractory Chronic Pain. *Front Comput Neurosci* **12**, 18 (2018).

90. Dworkin, R. H. *et al*. Validation of the Short-Form McGill Pain Questionnaire-2 (SF-MPQ-2) in Acute Low Back Pain. *The Journal of Pain* **16**, 357–366 (2015).

91. Deuschl, G. *et al*. A Randomized Trial of Deep-Brain Stimulation for Parkinson's Disease. *New England Journal of Medicine* **355**, 896–908 (2006).

92. Benabid, A. L. *et al*. Long-term suppression of tremor by chronic stimulation of the ventral intermediate thalamic nucleus. *The Lancet* **337**, 403–406 (1991).

93. Mayberg, H. S. *et al*. Deep Brain Stimulation for Treatment-Resistant Depression. *Neuron* **45**, 651–660 (2005).

94. Bergfeld, I. O. *et al*. Deep Brain Stimulation of the Ventral Anterior Limb of the Internal Capsule for Treatment-Resistant Depression: A Randomized Clinical Trial. *JAMA Psychiatry* **73**, 456–464 (2016).

95. Mosley, P. E. *et al*. A randomised, double-blind, sham-controlled trial of deep brain stimulation of the bed nucleus of the stria terminalis for treatment-resistant obsessive-compulsive disorder. *Transl Psychiatry* **11**, 1–17 (2021).

96. Holtzheimer, P. E. & Mayberg, H. S. Deep Brain Stimulation for Psychiatric Disorders. *Annu Rev Neurosci* **34**, 289–307 (2011).

97. Boon, P. *et al*. Deep Brain Stimulation in Patients with Refractory Temporal Lobe Epilepsy. *Epilepsia* **48**, 1551–1560 (2007).

98. Feldmann, L. K. *et al*. Toward therapeutic electrophysiology: beta-band suppression as a biomarker in chronic local field potential recordings. *npj Parkinsons Dis.* **8**, 1–9 (2022).

99. Bradley, C., Nydam, A. S., Dux, P. E. & Mattingley, J. B. State-dependent effects of neural stimulation on brain function and cognition. *Nat Rev Neurosci* **23**, 459–475 (2022).

100. Van Hartevelt, T. J. *et al*. Neural Plasticity in Human Brain Connectivity: The Effects of Long Term Deep Brain Stimulation of the Subthalamic Nucleus in Parkinson's Disease. *PLoS ONE* **9**, e86496 (2014).

101. Avitan, L. & Stringer, C. Not so spontaneous: Multi-dimensional representations of behaviors and context in sensory areas. *Neuron* **110**, 3064–3075 (2022).

102. West, J. *et al*. Machine learning seizure prediction: one problematic but accepted practice. *J. Neural Eng.* **20**, 016008 (2023).

103. Westfall, P. H. & Young, S. S. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. (John Wiley & Sons, 1993).

104. Sani, O. G. *et al*. Mood variations decoded from multi-site intracranial human brain activity. *Nat Biotechnol* **36**, 954–961 (2018).

105. Widge, A. S. *et al*. Treating refractory mental illness with closed-loop brain stimulation: Progress towards a patient-specific transdiagnostic approach. *Experimental Neurology* **287**, 461–472 (2017).

106. Dougherty, D. D. *et al*. A Randomized Sham-Controlled Trial of Deep Brain Stimulation of the Ventral Capsule/Ventral Striatum for Chronic Treatment-Resistant Depression. *Biological Psychiatry* **78**, 240–248 (2015).

107. Levinson, L. H. *et al*. (in preparation) Number of CCEP Trials Influences Quantified Variability.

108. Miller, K. J., Müller, K.-R. & Hermes, D. Basis profile curve identification to understand electrical stimulation effects in human brain networks. *PLoS Comput Biol* **17**, e1008710 (2021).

109. Huang, H. *et al*. Electrical Stimulation of Temporal and Limbic Circuitry Produces Distinct Responses in Human Ventral Temporal Cortex. *J. Neurosci*. **43**, 4434–4447 (2023).

110. Habets, J. G. V. *et al*. An update on adaptive deep brain stimulation in Parkinson's disease. *Movement Disorders* **33**, 1834–1843 (2018).

111. Mayberg, H. S. Electrophysiological Biomarkers to Optimize DBS for Depression. (2018).

112. Levinson, L. H. *et al*. Intraoperative Characterization of Subthalamic Nucleus-to-Cortex Evoked Potentials in Parkinson's Disease Deep Brain Stimulation. *Front. Hum. Neurosci*. **15**, (2021).

113. Sisterson, N. D. *et al*. Electrocorticography During Deep Brain Stimulation Surgery: Safety Experience From 4 Centers Within the National Institute of Neurological Disorders and Stroke Research Opportunities in Human Consortium. *Neurosurgery* **88**, E420 (2021).

114. Mercier, M. R. *et al*. Advances in human intracranial electroencephalography research, guidelines and good practices. *NeuroImage* **260**, 119438 (2022).

115. Bryan, M. J., Jiang, L. P. & Rao, R. P. N. Neural Co-Processors for Restoring Brain Function: Results from a Cortical Model of Grasping. Preprint at https://doi.org/10.48550/arXiv.2210.11478 (2023).

116. Rao, R. P. Towards neural co-processors for the brain: combining decoding and encoding in brain-computer interfaces. *Curr Opin Neurobiol* **55**, 142–151 (2019).

117. Rao, R. P. N. *Brain Co-Processors: Using AI to Restore and Augment Brain Function*. http://arxiv.org/abs/2012.03378 (2020) doi:10.48550/arXiv.2012.03378.

118. Klein, E., Brown, T., Sample, M., Truitt, A. R. & Goering, S. Engineering the Brain: Ethical Issues and the Introduction of Neural Devices. *Hastings Center Report* **45**, 26–35 (2015).

119. Yuste, R. *et al*. Four ethical priorities for neurotechnologies and AI. *Nature* **551**, 159–163 (2017).

120. Schönau, A. *et al*. Mapping the Dimensions of Agency. *AJOB Neuroscience* **12**, 172–186 (2021).

121. Klein, E. *et al*. Brain-computer interface-based control of closed-loop brain stimulation: attitudes and ethical considerations. *Brain–Computer Interfaces* **3**, 140–148 (2016).

122. de Haan, S., Rietveld, E., Stokhof, M. & Denys, D. Effects of Deep Brain Stimulation on the Lived Experience of Obsessive-Compulsive Disorder Patients: In-Depth Interviews with 18 Patients. *PLoS One* **10**, e0135524 (2015).