UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

PARAMETERIZING THE RELATIONSHIP BETWEEN EEG AND EMG DURING PRESCRIBED MOVEMENT

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

ROBEL TESFASELASSIE

Norman, Oklahoma

2025

PARAMETERIZING THE RELATIONSHIP BETWEEN EEG AND EMG DURING PRESCRIBED MOVEMENT

A THESIS APPROVED FOR THE SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Chad Roller, Chair

Dr. Hazem Refai, Co-Chair

Dr. Joseph Havlicek

Dr. Samuel Cheng

Dedication

To Papa and Mama, whose unconditional love and guidance I treasure so dearly,

To Feven, for being my rock, my boba buddy, and the best sister I could ever ask for,

To my heroes in the field of healthcare, who impact the world in ways I hope to emulate,

And to Ace and Angel, my joyful, faithful, four-legged companions,

This work is dedicated to you.

"Unless the Lord builds the house, the builders labor in vain."
Psalms 127:1a

Acknowledgements

First, I would like to extend my gratitude and appreciation to my advisors Dr. Hazem Refai and Dr. Chad Roller for their invaluable support and guidance in the development of this thesis and related research. In addition, I would like to thank the distinguished members of my committee, Dr. Joseph Havlicek and Dr. Samuel Cheng, for their time and effort in reviewing this thesis. Many thanks to Ms. Denise Davis for her ever-present support throughout my graduate research experience. Thank you to the Sooner Family, at both OU-Norman and OU-Tulsa, for developing me into the engineer I am today. Finally, thank you to my family and my friends, as well as my fellow graduate students, for being in my corner as I've embarked on this academic journey.

Table of Contents

Abstract

Understanding the relationship between electrical activity within the cerebral cortex of the brain and muscular output is essential for advancing neuroprosthetic design and neuromotor rehabilitation. This thesis establishes and parameterizes the relationship between electroencephalography (EEG) and electromyography (EMG) signals during the prescribed contraction of the forearm and hand muscles in squeezing motions. First, it was confirmed that a reduction in alpha-band activity at the C3 EEG electrode corresponds to increased signal power in the flexor carpi radialis EMG electrode, supporting the hypothesis of cortical-motor coupling. Second, autoregressive (AR) models were constructed to characterize the spectral and stochastic properties of both EEG and EMG signals, achieving spectral flatness values of 0.91 and 0.96 in the EEG and EMG models, respectively. Finally, a novel transfer function is proposed that defines the EEG-to-EMG system - yielding EMG-like output with a normalized spectral error of approximately 0.37 - and two methods of transfer function order selection are compared. These findings demonstrate the utility of system identification techniques in mapping cortical activity to muscular response, contributing to the development of interpretable and straightforward EEG-EMG models. Further, it may help to facilitate real-time control in brain-computer interfaces and related assistive technologies.

1 Introduction

A particularly potent concept was developed in the early 1970s to make day-to-day activities more accessible and autonomous for individuals with severe musculoskeletal paralysis. Termed "Brain Computer Interface" (BCI) by Jacques Vidal in 1973, the soon-to-come technology would ideally read the user's mind, decode movement intent, and follow those instructions to control a prosthetic device. Over five decades later, ready-to-use BCIs are being produced by companies such as Synchron, Neuralink, and Blackrock Neurotech. For their users, the impact has been nothing short of "world-changing".

Many currently used BCIs are invasive, meaning that the device is implanted onto the surface of the brain itself. Although this excellent proximity to relevant electrophysical signals results in clear and detailed signals, invasive BCIs are prohibitively expensive, involve risky surgical procedures, and have hitherto not been received well by the general public (Pew Research, 2022). Thus, a lot of recent research has been focused on the development of decoding algorithms for *non-invasive* BCIs that can decode the somewhat noisy neural signals collected from a vantage point *outside* the skull using an electroencephalogram (EEG). In alignment with this broader trend, the research in this thesis adopts a non-invasive approach by utilizing EEG signals recorded from the scalp.

Existing non-invasive BCI algorithms have focused on complex machine learning approaches, such as neural networks or ensemble models, that classify movements into predefined categories of movement, such as "move hand left" or "move hand right". Other approaches attempt to use EEG data to predict degrees of motion or force, such as the power of a throw, or the degree of a bent elbow. However, current research has not fully explored the

possibility of using EEG data to approximate electric signals in the forearm muscles, measured with an electromyogram (EMG). This approach not only allows for a more precise control of grasping motions in BCI applications, but it also could provide a clearer, non-invasive understanding of the system that electrical signals go through as they propagate through the central and peripheral nervous system networks, from brain to arm.

In this thesis, this approach was explored through the development of an EEG-EMG transfer function, based on individual stochastic models for EEG and EMG data. Using stochastic models to develop this relationship potentially adds to the current understanding of these noisy electrophysiological data sources, and it accounts for the inherent noise in these non-invasive data collection processes. Finally, included in this thesis is an analysis of the EEG and EMG data used in this research, adding to the current understanding of electrical activity in the nervous system as it relates to voluntary movement.

1.1 Thesis Contributions

This thesis was developed with the objective of parameterizing the relationship between EEG activity and EMG activity during movement. Encompassing linear and stochastic methods, the research detailed in this study enhances the current understanding of the electrophysiological link between electrical activity in the motor cortex of the brain and electrical activity in the forearm muscles during movement, while also providing a possible basis on which to build future algorithms for non-invasive, precise-control BCIs.

The main contributions of this thesis are summarized below:

[1] Confirm that a decrease in EEG alpha band activity in the C3-EEG electrode is related to an amplitude increase in EMG data at the flexor carpi radialis electrode during prescribed movement.

[2] Develop autoregressive models for the C3 electrode in EEG data and the flexor carpi radialis electrode in EMG data during prescribed movement.

[3] Devise a novel transfer function that describes the relationship between the EEG data and EMG data during prescribed movement.

[4] Evaluate EEG-EMG transfer functions using normalized spectral error.

1.2 Content Summary

Consistent with the outlined objectives, this thesis is divided into six chapters summarized here. Chapter 1 covers the objectives of this work and a high-level overview of the research and analyses conducted in this thesis. Chapter 2 summarizes the background and related works in this field of research. Chapter 3 outlines the data collection systems used for this study and covers preliminary data analysis. Chapter 4 describes the development and evaluation of individual autoregressive models that best characterize EEG data from the C3 electrode and EMG data from the flexor carpi radialis electrode during prescribed movement. Chapter 5 works through the development and evaluation of a transfer function that relates EEG data to EMG data during voluntary movement using the stochastic models developed in Chapter 4. Chapter 6

provides a summary of the work presented in this thesis and discusses possible directions for further exploration in this field.

2 Background

Electrophysiological activity has long been understood to be a noisy random process, especially as recorded from the surface of one's scalp and body as in the case of the electroencephalogram (EEG) and the surface-electromyogram (sEMG), respectively. In BCI applications, this is the driving force behind the challenge of decoding intent from noisy, non-deterministic brain signals. Despite recent advancements in artificial intelligence and machine learning, decoding movement intent from noisy EEG data remains challenging and requires extremely complex algorithms. Translating movement-related EEG waveforms into fine-motor movement commands presents even more of a challenge.

At its core, the task of decoding movement intent is centered around elucidating the link between brain activity and modes of movement. In simple cases like unilateral grasping motions, movement onset by way of muscle contraction is often indicated by an increase in the EMG waveform amplitude. Movement onset may not be as clear in more complex cases, but EMG data typically changes in some meaningful way to reflect salient properties of muscle recruitment. In both of these cases, the task of decoding movement intent from EEG data can be reframed as defining the relationship between EEG data and EMG data. Using this new problem statement, defining this relationship will require a characterization of the system that electrical signals travel through as they propagate from the motor cortex to the peripheral nervous system network in the limbs.

To date, studies have explored the EEG-EMG relationship in a variety of approaches that are mostly centered around coupling methods. In one study, researchers used both linear and non-linear methods to evaluate corticomuscular coupling during movement, and they observed

high EEG-EMG coherence (linear coupling) in beta/gamma bands during grasping movements and motor intention, while also observing higher mutual information (nonlinear coupling) during movements/motor intention than during regular motor imagery or resting states (Tun, 2018). This paper also showed higher contralateral coupling than ipsilateral, reinforcing the understanding that the brain activity related to a unilateral movement is focused on the side of the motor cortex opposite the side of the physical movement. Also, in many other studies, corticomuscular coherence was shown to be higher for healthy subjects than for stroke patients and those with neuromuscular diseases like ALS (Liu, 2019). Another study used a novel method called variable scale symbolic transfer entropy (VS-STE) to compare corticomuscular coupling in post-stroke patients and healthy controls (Gao, 2018). Interestingly, it was found that post-stroke patients had higher VS-STE values on the affected side than on the unaffected side, and generally post-stroke patients had higher corticomuscular coupling than healthy patients. This result suggests that VS-STE coupling could reflect the effort exerted on the brain-muscle pathway, since the impacted corticomuscular pathways in post-stroke patients are often impacted significantly and require more training to rehabilitate and use. Clearly, corticomuscular coupling has proved very useful in determining valuable information about the brain-muscle connection in voluntary movement contexts. Studies have even used EEG-EMG coherence levels to classify combined EEG and EMG datasets into grasping tasks with different weights, with accuracy levels that vary from 58% to 94%, depending on where the EMG electrode is placed and the length of time spent grasping the object (Cisotto, 2018).

All-pole linear predictive modeling - a well-established method for representing correlated random processes - has also been explored as a way to decompose EEG and movement-related EMG waveforms into compact features. As early as 1970, Fenwick et al.

demonstrated the possibility of using autoregressive models to simulate EEG activity (Fenwick, 1970). Interestingly, their conclusion that 20–30 second epochs result in stable, accurate AR parameter estimation matches the epoch length chosen in this thesis. Since then, numerous studies have further explored the use of EEG AR parameters for various applications. One such study was able to classify between patients with Parkinson's disease and control patients with an 85% accuracy using just the poles determined from the linear predictive model of each patient's EEG data (Anjum, 2020). Similar studies have succeeded at compressing the spectral information in EMG data. Carotti et al. used a modified version of the algebraic code excited linear prediction algorithm (ACELP) - which is typically used for speech signal coding - to compress sEMG data during isometric bicep contraction (Carotti, 2007). This resulted in a compression ratio of 87% and a mean square error of 6.8% after reconstruction. Errors in the properties of the reconstructed signal like mean power spectral frequency and 3rd order power spectral moments were within 6%, showing minimal loss of information in linear predictive analysis-based compression. Evidently, linear predictive modeling has proven useful in characterizing EEG and EMG processes.

Although the aforementioned studies have proven useful in understanding the two signal sources and the dynamics of their relationship, very few studies have directly connected the two using a defined, parametric system. Having such a system could contribute to our understanding of the EEG-EMG relationship, as well as the simplification and fine-motor-precision of existing BCI technologies. To that end, this thesis defines such a system and transfer function using stochastic models that account for the random nature of electrophysiological signals in the brain and forearm muscles.

3 EEG/EMG Data Collection and Analysis

Provided in this chapter is an overview of the electrophysiological data collection systems used in this thesis research, as well as brief analysis on the data collected. This section gives the reader some insight into the nature of the data that will be dealt with in subsequent chapters.

3.1 Data Collection Paradigm

The experiment central to this thesis was conducted in the Neuro Lab at the University of Oklahoma – Tulsa campus. The IRB number for this study is #17479. The participant was asked to sit down and hold a standard hand-grip dynamometer (also known as a grip-strength tester) in their right hand. Over the course of a few minutes, the participant was instructed to alternate between squeezing their right hand (at 80% of maximum power) for 30 seconds, and relaxing their right hand for 40-50 seconds. As the participant engaged in these tasks, researchers used a 64-channel EEG and a 4-channel EMG to simultaneously collect electrophysiological data at two locations in the body: the scalp, and the forearm. This process was repeated five times, resulting in five datasets of simultaneously-collected EEG and EMG data.

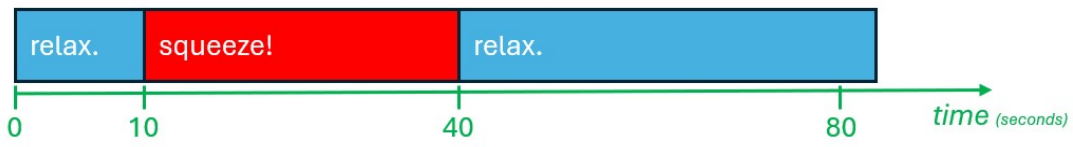


***Figure 3-1**. Timing scheme of the data collection paradigm.*

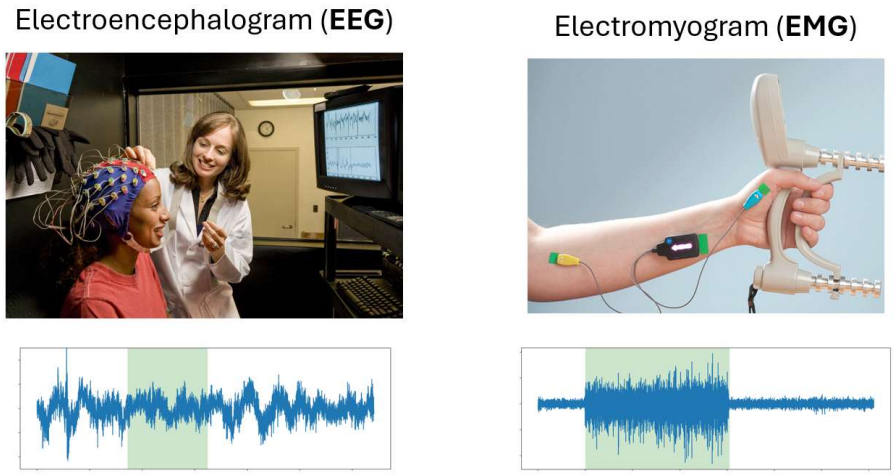


***Figure 3-2**. Top) Examples of EEGs and EMGs in use. Bottom) Examples of their respective waveforms. The region highlighted in green indicates a squeezing period.*

3.2 EMG Data Collection

3.2.1 EMG Data Collection Equipment

The EMG sensor used in this research is the wireless 4-channel Delsys Trigno Quattro Sensor (Delsys Inc, 2017).

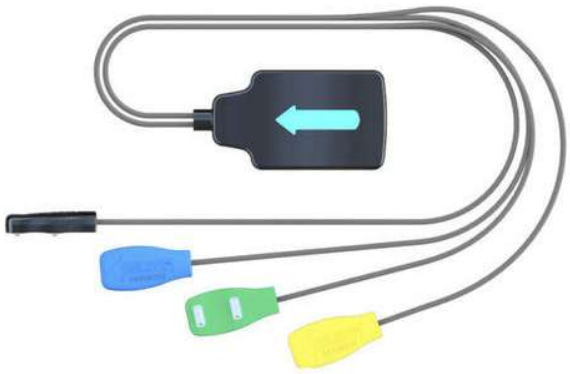


***Figure 3-3**. The Delsys Trigno Quattro sensor.*

Equipped with a built-in bandpass filter, the manufacturer-defined bandwidth of the data collected using this sensor is 10-850Hz. Given that the sampling rate of the Quattro sensor is 2222.2 Hz, this bandwidth is within the Nyquist limit and ensures the absence of aliasing effects while also removing common biomedical artifacts like unrelated movement (< 5 Hz), respiration (0.1-0.5Hz), or heartbeat (1-2Hz).

The four electrodes of the EMG were placed on the following places on the participants' right arm: the triceps brachii (electrode #1), the biceps brachii (electrode #2), the flexor carpi ulnaris (#3), and the flexor carpi radialis (electrode #4).

The Quattro sensor connects to the EMGWorks application, which receives and displays EMG signals in real time. After collection, the data is saved as a Delsys proprietary .shpf file, which is later converted into a .csv file using the application's export function. The .csv file format works well with Python through the pandas library (Python Data Analysis), which allows the user to easily import data stored in the .csv format as an array.

3.2.2 EMG Data Preprocessing

The data preprocessing pipeline for EMG data, as shown in Figure 3-4, starts with loading the data into Python from the EMGWorks .csv file. Since the data is collected in millivolts, the next step is to convert the data into volts for unit consistency.

***Figure 3-4**. EMG data preprocessing pipeline.*

Although it is not the primary source of hand-squeezing force, it has been shown that activation of the flexor carpi radialis muscle, which stabilizes the wrist and hand, correlates well with grip force (Agarwal, 2018). Therefore, of the four EMG electrodes, the electrode placed on the flexor carpi radialis (#4) was determined to be the most useful electrode in measuring electrophysiological activity related to the squeezing task. Additionally, good placement with little adjacent muscle interference is easy to ensure given the flexor carpi radialis muscle's width and surface-level position. From here on, discussions regarding EMG data refer to the electrophysiological data collected at the flexor carpi radialis.

As mentioned earlier, the data collection equipment has a built-in bandpass filter with a 10-850 Hz bandwidth. To further ensure that relevant spectral information is retained and noise is eliminated, a digital fifth-order Butterworth bandpass filter is also implemented, with cutoff frequencies 10-150 Hz. The upper cutoff frequency reflects the accepted consensus that the dominant spectral energy in sEMG signals is below 150 Hz (De Luca, 2002).

Finally, the EMG data is downsampled from 2.222 kHz to 1 kHz. Later analysis involves both EEG and EMG data, so it is crucial that both data types have the same sampling rate. The 1kHz sampling rate was chosen because it fulfills the Nyquist requirement for both EEG and

EMG data (relevant frequencies do not exceed 150 Hz) without excessive oversampling. Because both EEG and EMG data have the same sampling rate, the onset and end of the squeeze in the EMG data can more easily be marked with the help of the synchronization markers in the EEG data.

3.3 EMG Data Analysis

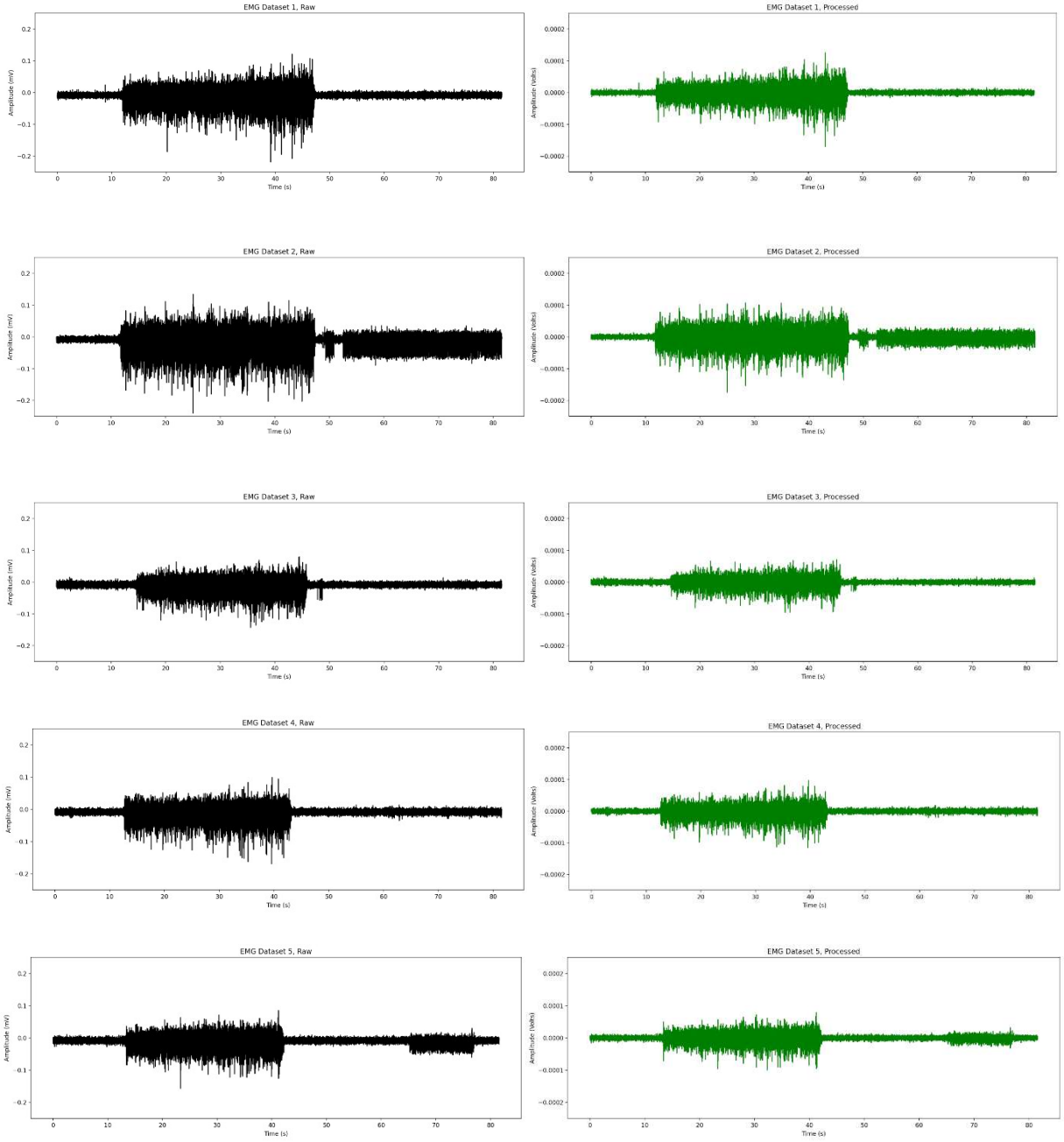


***Figure 3-5**. EMG data before (left) and after (right) preprocessing steps.*

After preprocessing steps, the filtered EMG data has a reduction in overall power as frequencies above 150 Hz have been attenuated. Aside from the exact timing of a few voltage spikes (which have also been altered by the resampling step), the data retains its original shape, and it is clear that muscle activation is represented by an increase in the amplitude of the EMG waveform. An example of this clear demarcation can be seen in Figure 3-6. In each of the datasets, the squeeze occurs between t = 10 seconds and t = 50 seconds. The exact duration of the squeeze varies from dataset to dataset, ranging from 28 to 37 seconds.

In some of the datasets (especially datasets 2 and 5), significant increases in amplitude can be observed outside of the main squeeze period. It can be inferred that these regions are related to accidental muscle activation during the non-squeeze/rest period. Therefore, because it is possible that the rest period is contaminated by movement-related artifacts and noise, it is difficult to properly compare the non-squeeze EMG data to squeeze-related EMG data,. Further analysis and signal modeling is limited to the squeeze-region of the EMG data.

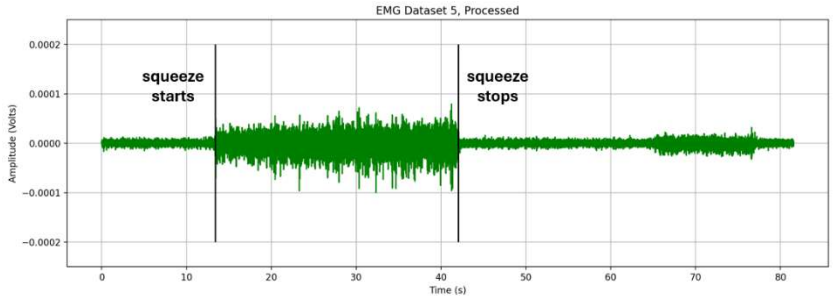


***Figure 3-6**. EMG dataset 5 with squeeze-start and squeeze-stop points marked.*

The distributions of the EMG datasets each have a clear zero mean, and each dataset has a variance between 2e-10 V^2 and 6e-10 V^2. However, the distribution of each EMG dataset is skewed slightly to the left, as shown by the median-mean relationship in Figure 3-7. This asymmetry in the collected and processed EMG data suggests that the data is not perfectly Gaussian. This skewness, which indicates that negative spikes are common in this study's EMG data, has also been shown in other studies like Nazarpour et. al (2013).

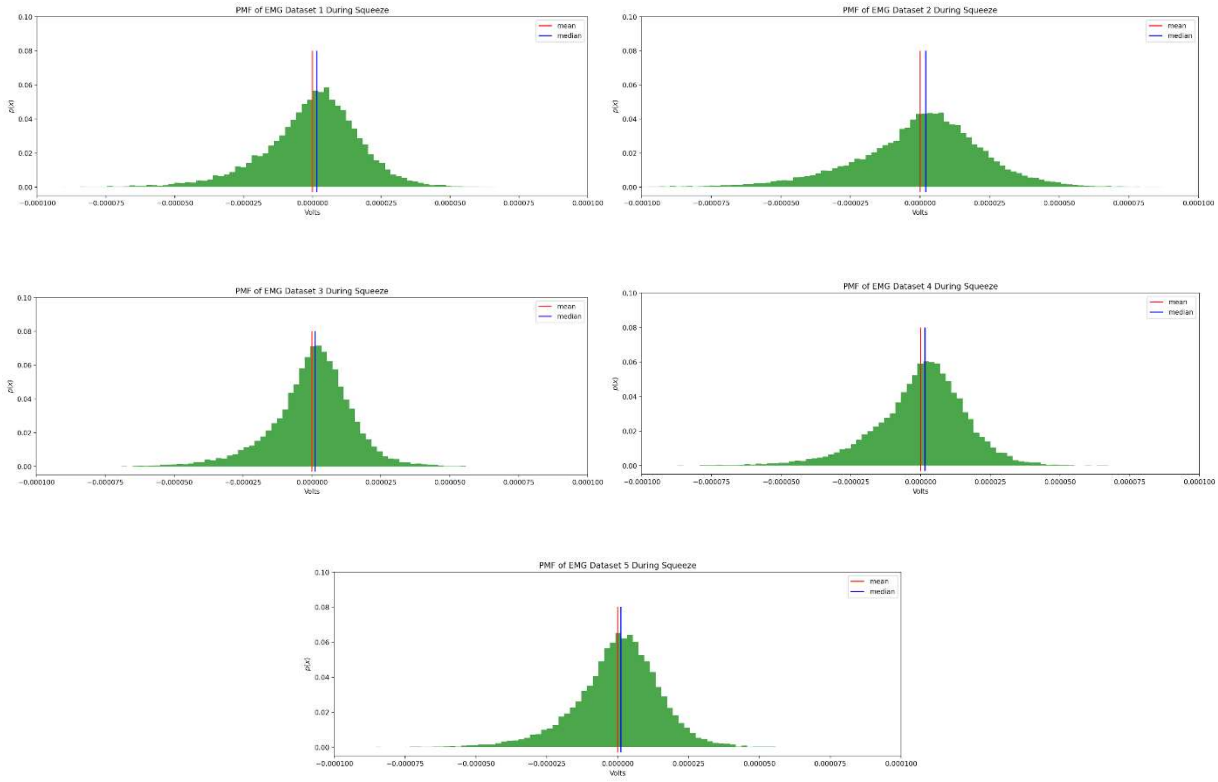


Figure 3-7*. Probability mass functions of EMG datasets, with means and medians marked in red and blue, respectively.*

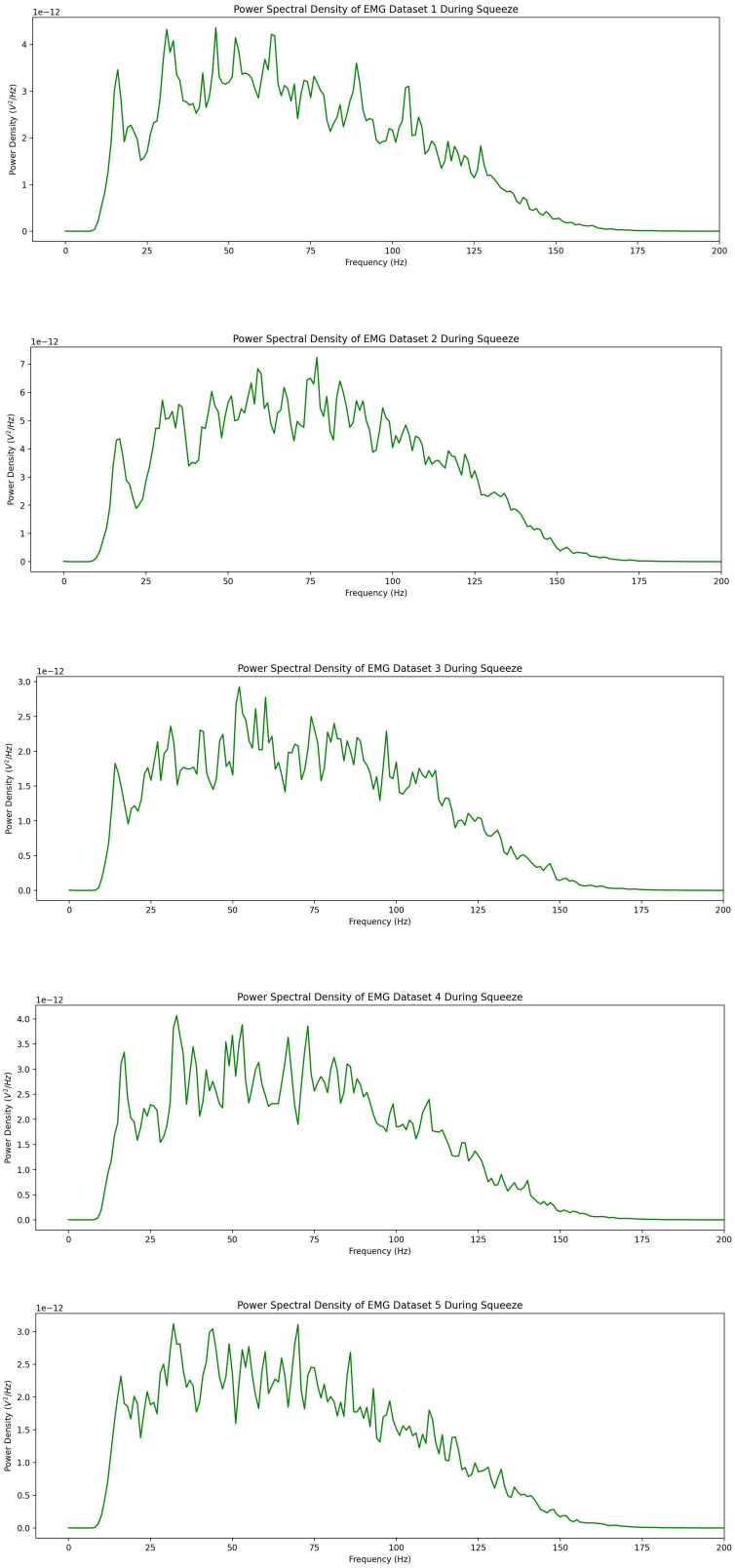


***Figure 3-8**. Power spectral density (using Welch's method) of processed EMG data.*

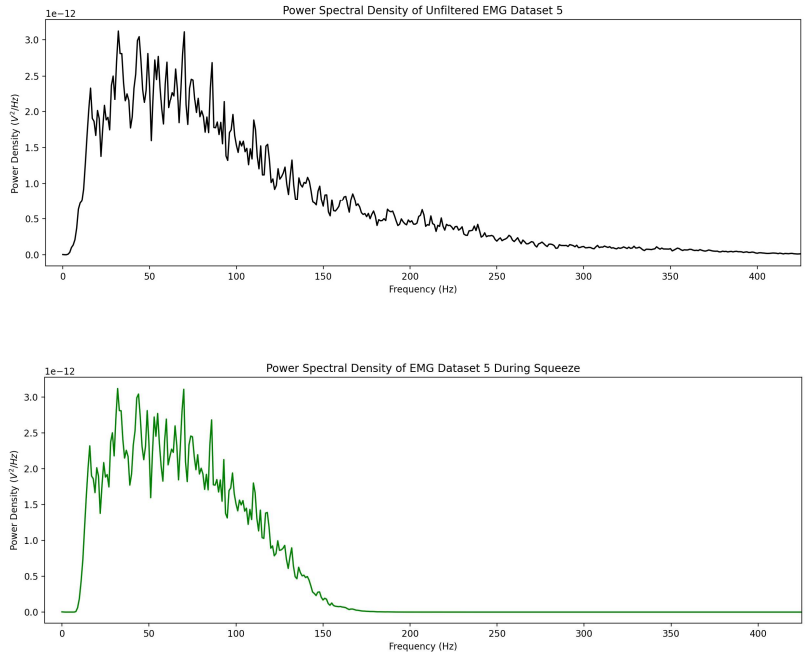


Figure 3-9. *Illustration of changes in the frequency spectrum after preprocessing.*

In the frequency domain, most of the power in EMG data is contained between 10 and 150 Hz. As seen in Figure 3-9, unfiltered EMG data contains a long tail of steadily decreasing power in frequencies above 150 Hz.

With some exceptions, the power in EMG data is relatively uniform between 25 and 80 Hz, and a slow decrease in power is observed as frequencies approach 150 Hz. Although seemingly random sharp peaks in power are observed at some frequencies, as seen in Figure 3-8, the exact location of these peaks change from dataset to dataset. The overall smooth trend of the power spectral density (PSD) in EMG data remains consistent across the EMG datasets.

To analyze how EMG power spectral density changes over time, spectrograms were computed for each dataset using a 1024 ms window and 75% overlap. As with the time-domain representation of EMG signals, broadband increases in power clearly correspond to periods of

muscle activation. During the squeeze period, the power is relatively uniformly distributed across the 25–100 Hz range. In some cases, a slight increase in EMG power can be observed at the end of the squeeze, likely signaling slight fatigue. In contrast, during non-squeeze periods, power is concentrated between 10–40 Hz and falls off sharply above this range. Additionally, the shape of the PSD during a resting period doesn't change much with time, aside from noise related to accidental movement. These differences in power distribution in the frequency domain could provide a useful means of characterizing and distinguishing between active and resting muscle states.

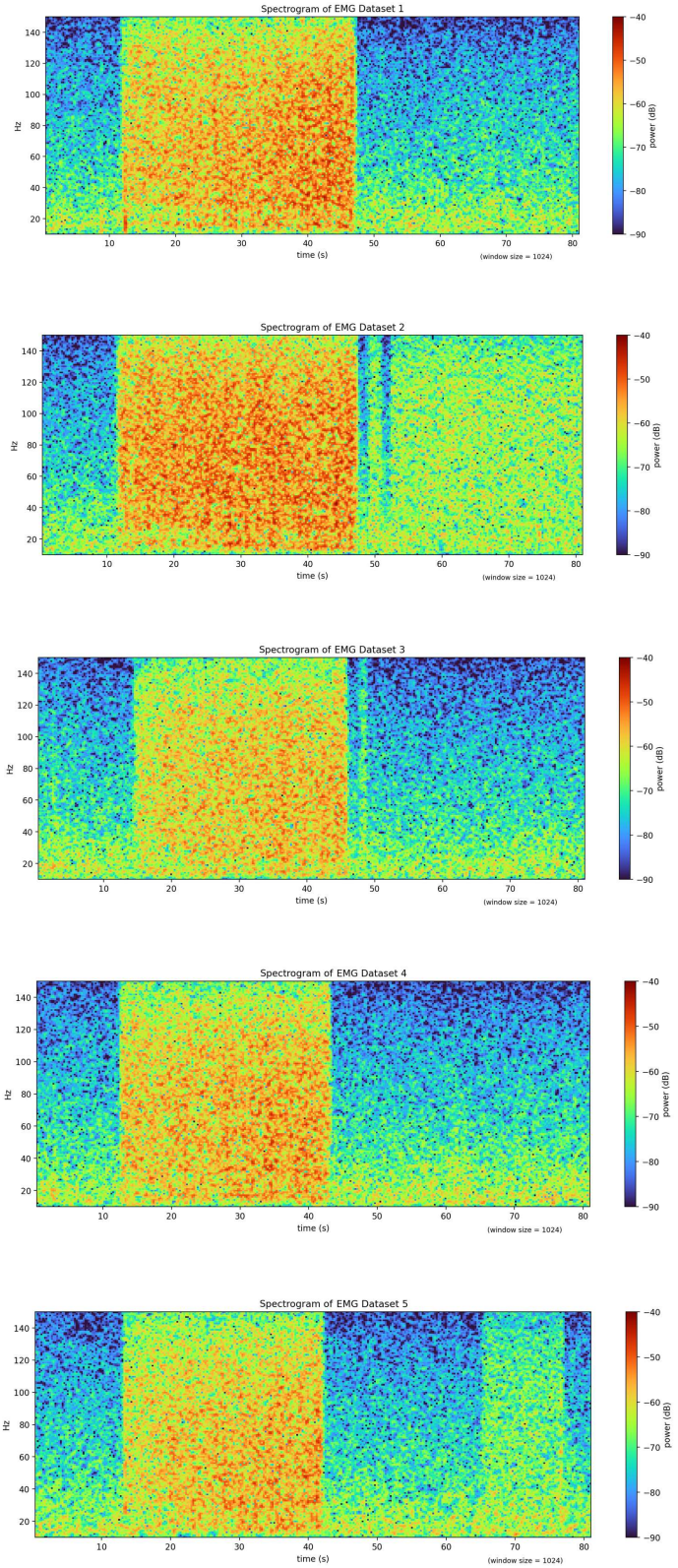


***Figure 3-10**. Spectrograms of EMG datasets using a 1024ms window and 75% overlap.*

KPSS test (Kwiatkowski–Phillips–Schmidt–Shin)		ADF test (augmented Dickey–Fuller)	
Results of KPSS Test:		**Results of Dickey-Fuller Test:**	
Test Statistic:	0.000907	Test Statistic:	-68.822597
p-value:	0.10	p-value:	0.000000
Lags Used:	144	Lags Used:	65.0
Critical Value (10%):	0.347	# Observations Used:	81510
Critical Value (5%):	0.463	Critical Value (1%):	-3.430430
Critical Value (2.5%):	0.574	Critical Value (5%):	-2.861575
Critical Value (1%):	0.739	Critical Value (10%):	-2.566789
= stationary		= stationary	

KPSS says: / ADF says:	stationary	non-stationary
stationary	**stationary**	**difference stationary**
non-stationary	**trend stationary**	**not stationary**

Figure 3-11*. Top) Results of KPSS and ADF stationarity tests on squeeze-related data in EMG dataset 5. Bottom) Interpretation of joint KPSS/ADF tests.*

Accurate characterization of a random process highly depends on its stationarity. The tendency of EMG data to slightly increase in power at the end of the squeeze raises questions about whether these EMG datasets are stationary. As seen in Figure 3-11, stationarity was confirmed using the KPSS and ADF tests. The KPSS test has a null hypothesis that the EMG dataset is stationary, which is not rejected, and the ADF test has a null hypothesis that the EMG data is not stationary, which is rejected.

Random processes can also be characterized by their autocorrelation functions, which tell the observer how the value of a signal at a given time-point is related to the values of past samples. Autocorrelation functions that quickly converge to zero indicate that the signal has little

dependence on previous values. As Figure 3-12 demonstrates, the estimated autocorrelation function of EMG data during the squeeze period converges to zero within 25-30 ms of lag, indicating that EMG data is correlated over 25 ms time-scales, with little dependence beyond this lag range.

It is clear from the top plot in Figure 3-12 (as well as in the plots of the EMG data in frequency and in time) that the EMG datasets have varying power levels. Since the amplitude envelope of EMG data can generally be understood to positively correlate with the force of muscle contraction, it is possible that this discrepancy stems from a slightly different force being exerted in each squeeze.

To allow for shape comparison independent of signal power, autocorrelation function estimates were normalized by dividing each dataset's autocorrelation estimate by the variance of their data. From the normalized autocorrelation function plot in the bottom of Figure 3-12, a pattern can be seen within the first 10-15 ms of lag. A consistent sharp dip from lag 0 to lag 5 is observed in all datasets, followed by a smaller rise and an even smaller dip between 5 and 15 ms. With small differences in the exact shape, each of the datasets have autocorrelation estimates that follow this pattern. Past 15-20 ms of lag, the EMG datasets converge to near-zero values in a less consistent manner.

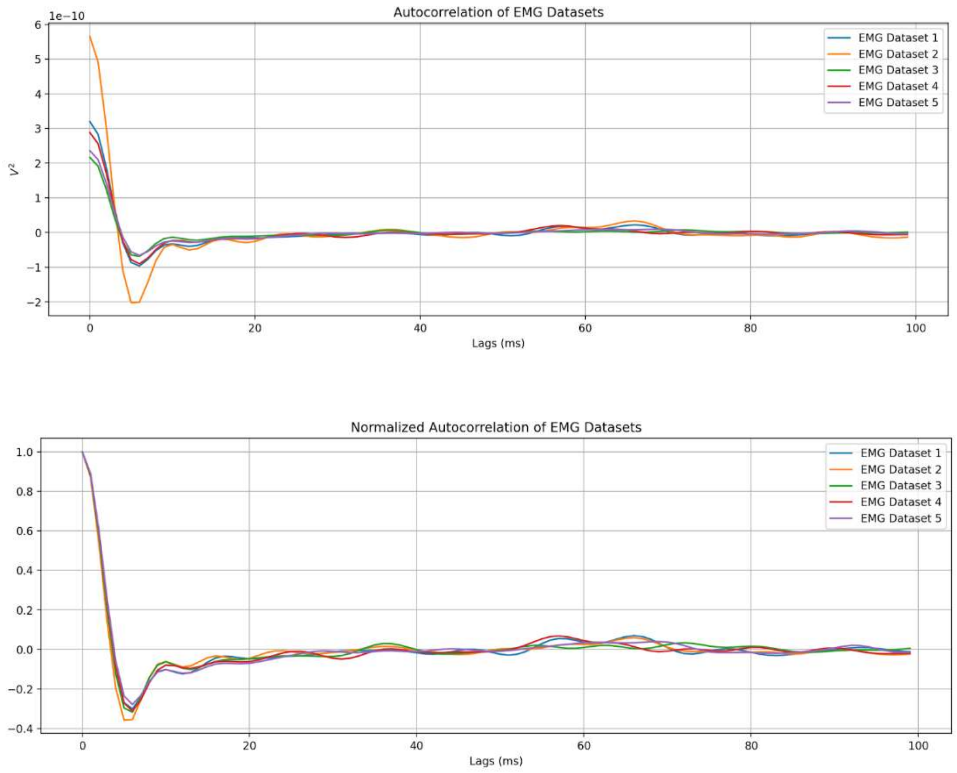


***Figure 3-12**. Top) Autocorrelation function estimates of EMG datasets. Bottom) Plot of normalized autocorrelation.*

3.4 EEG Data Collection

3.4.1 EEG Data Collection Equipment

The main EEG data acquisition instruments used in this study are the 64-channel BrainVision actiCAP Snap (Brain Products, 2019a) and the actiCHAMP PLUS amplifier (Brain Products, 2019b).

Figure 3-13. *Left) BrainVision actiCAP Snap EEG system. Right) BrainVision actiCHAMP PLUS amplifier.*

Before EEG data collection begins, it is crucial to check that all skin-electrode impedances are low, which helps promote good signal quality. This involves inserting conductive solution/gel in between the electrodes and the scalp of the participant until impedances are below a threshold of 20 kilohm.

As electrodes pick up extremely low-amplitude electrical signals from the brain, the signals are amplified by the actiCHAMP PLUS amplifier, which also uses a timing module to help synchronize each of the 64 signals with each other and with the timing of the squeeze-hand and relax-hand commands. This step is extremely important, as the EEG data needs to have built-in stimulus-onset indicators at the correct times. Also included in the amplifier's data processing is a 60 Hz notch filter to remove electrical line noise from surrounding power sources.

After the signals are synchronized and amplified, the data is loaded into the BrainVision Analyzer 2.0 desktop application, which can be used to view the data that has been synchronized with the stimuli onsets. The Analyzer 2.0 GUI also allows the user to easily export data along with two metadata files (.vhdr and .vmrk). The main data file format, .eeg, can only be interacted with using either the MNE-Python library, or the EEGLAB (MATLAB) library. In this case, preprocessing steps were carried out in Python, using the MNE-Python library, the NumPy library, and the SciPy library.

3.4.2 EEG Data Preprocessing

The collected EEG data goes through a standardized data preprocessing pipeline before further calculations and analysis are performed. There are three main steps of EEG data preprocessing.

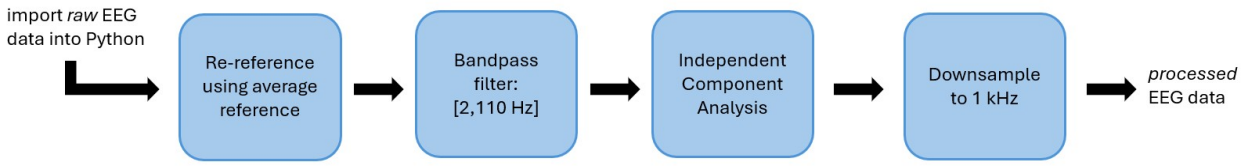


Figure 3-14*. EEG data preprocessing pipeline.*

a) Electrode Average Re-referencing

Each EEG channel records the voltage at a specific location on the scalp in reference to another location on the scalp. This reference location, or reference electrode, can often be very noisy. In addition, the choice of the reference electrode's location is subjective – some studies

use the Cz electrode, and other studies use FCz – and this ambiguity can make it hard to compare datasets and results. To account for this issue, a common solution is to create another artificial electrode by averaging all 64 EEG channels, and the electrodes are re-referenced with this artificial electrode as the "average" reference.

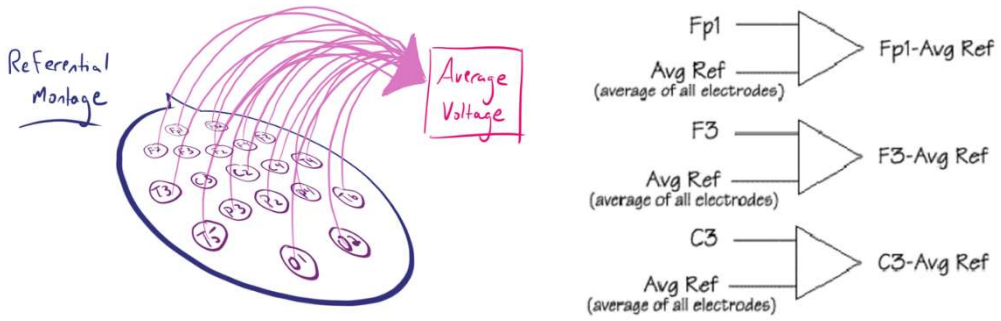


***Figure 3-15**. Illustration of the average re-referencing process.*

b) Bandpass filtering

Next, the re-referenced data is bandpass filtered. Quiet and subtle EEG data is often corrupted by loud and noisy non-brain data, known as artifacts. Examples of artifacts include subject movement, blinking, strong heartbeats, and respiration. Fortunately, most noise artifacts occupy frequencies that are well below that of relevant EEG data – for example, respiration occurs at 0.25 Hz and pulses typically hover around 1-2 Hz. Therefore, a lot of low frequency noise can simply be filtered out. In addition, bandpass filtering can be used to focus on a certain range of frequencies in EEG data. EEG oscillations above 110 Hz are not typically used in studies pertaining to motor functions, so those high-gamma frequencies can be filtered out. However, bandpass filtering requires tradeoffs. A frequency range of interest may occupy some of the same frequencies as some movement artifacts/noise. How much attenuation to apply in

each frequency region is up to the researcher; in this case, a bandpass filter of [2Hz, 110Hz] was used to sufficiently account for common artifact sources without removing too much relevant information in the lower ranges of the EEG spectrum.

c) Independent Component Analysis

The last step in the data preprocessing pipeline is Independent Component Analysis (ICA). This step helps remove some of the leftover noise that wasn't removed by the bandpass filtering step. ICA is a source separation technique that is used to separate noise sources from relevant signal sources. It does this by determining the most prominent spatial patterns associated with certain patterns in time. Each pattern returned is an "independent component". To the trained eye, it is typically quite clear which independent components are noise. For example, blinks can be pointed out as high-powered patterns that are located near the front of the head, next to the eyes.

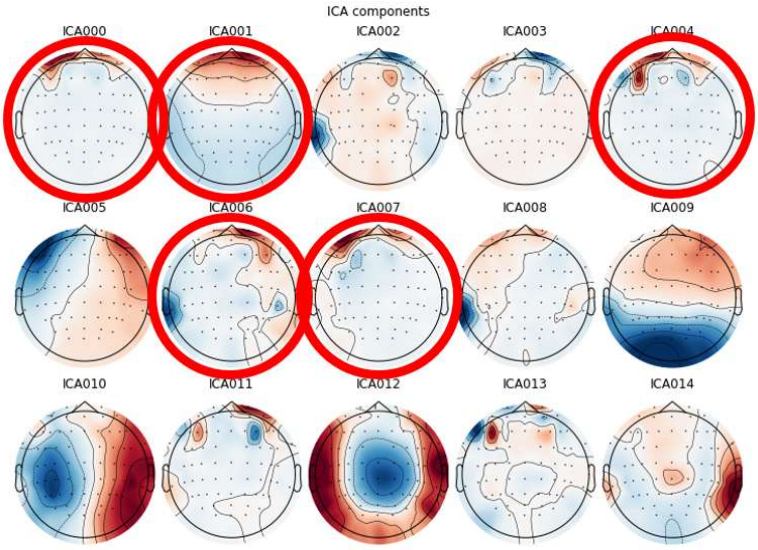


***Figure 3-16**. Example of the spatial components returned by ICA.*

The red circles indicate components that likely represent blinking-related noise. After the noise components are removed, the signal is reconstructed with the remaining components, and this step is complete. Finally, EEG data from all 64 channels is resampled from 5 kHz to 1 kHz to match the sampling rate of the EMG data.

The BrainVision actiCHAMP Snap EEG system recorded electrical data at 64 different places on the scalp. Since this study is focused on movement-based paradigms, the most relevant electrodes in this study are the electrodes C1, C3, C5, Cz, C2, C4, and C6 due to their position directly over the motor cortex portion of the brain.

Ultimately, due to the central location of the C3 electrode over the brain's left motor cortex, it was decided to focus efforts on finding a transfer function between electrode C3 and the flexor carpi radialis electrode. The choice of the left motor cortex stems from the fact that EEG-EMG coupling has been shown to be strongly contralateral (Tun, 2018).

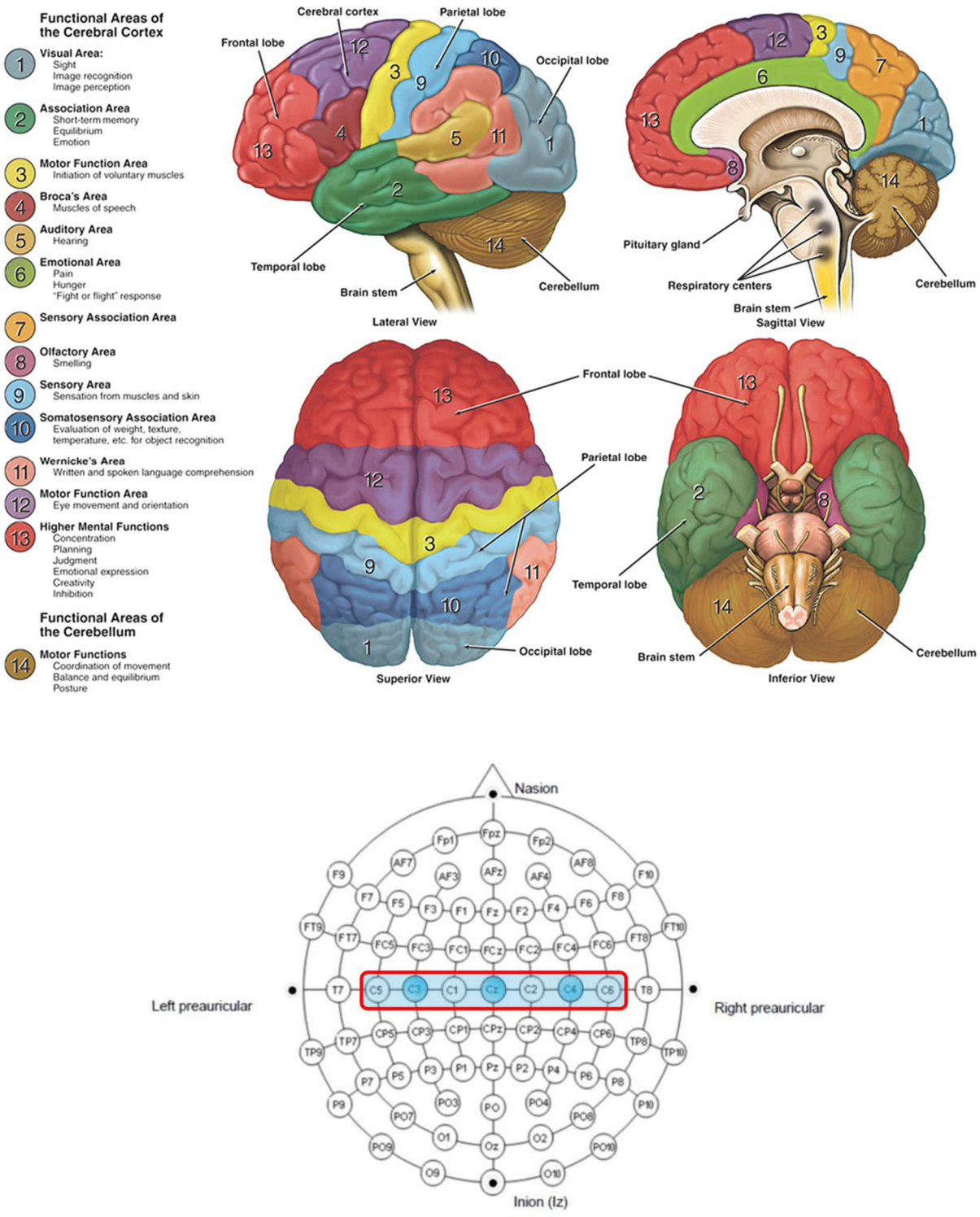


Figure 3-17*. Top) Functional areas of the brain. Bottom) 64-channel EEG map. Electrodes that are situated over the motor cortex region are highlighted.*

3.5 EEG Data Analysis

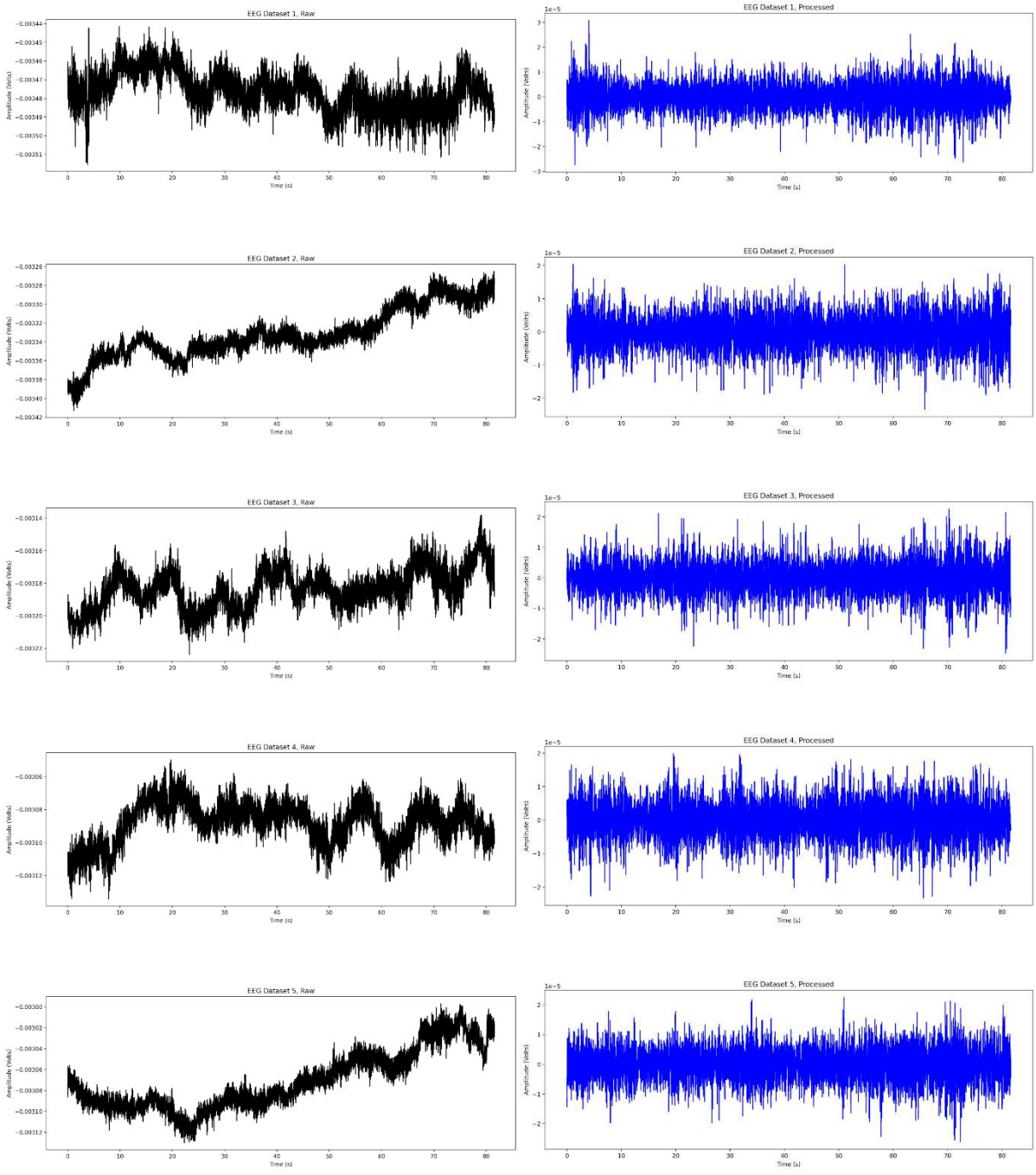


***Figure 3-18**. EEG data before (left) and after (right) preprocessing steps.*

Preprocessing steps for EEG data are extremely crucial, as noise from a variety of irrelevant sources can easily drown out subtle and low-amplitude brain signals. This can clearly be observed in Figure 3-18, as raw EEG data polluted with noise has amplitudes that are on a different order of magnitude as compared to processed EEG data. In addition, average re-referencing steps have helped zero-center the data. Finally, low-frequency noise from powerful interference like subject movement, blinking, heartbeat, and respiration are removed with the bandpass filter and ICA-based artifact removal. The result is a clean, smaller-amplitude dataset that still retains the envelopes and spikes that characterize the EEG data in the time domain. It is also clear that EEG data has much less power than EMG data, which is likely due in part to the larger amounts of attenuation experienced by brain signals as they propagate past the skull.

However, even after the discussed preprocessing steps, it is not trivial to identify the neural markers of voluntary movement from a simple visual inspection of time-series EEG data. To illustrate this, Figure 3-19 shows an EEG dataset with squeeze-start and squeeze-stop points marked in time. As with the EMG data analysis in section 3.3, the majority of the following EEG data analysis will be focused on the squeeze period.

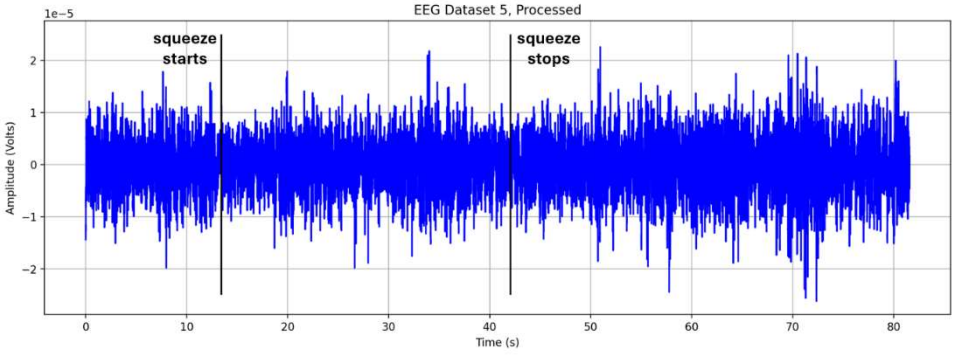


***Figure 3-19**. EEG dataset 5 with squeeze-start and squeeze-stop points marked.*

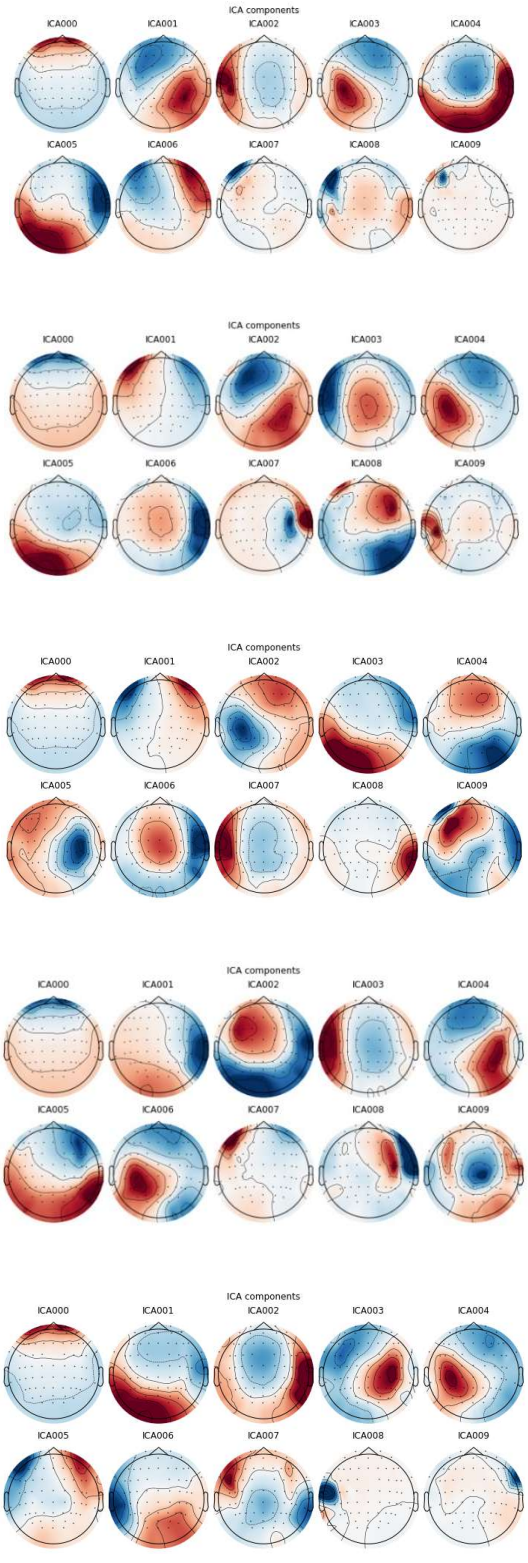


Figure 3-20. *Independent Component Analysis for EEG datasets 1-5, respectively. The most powerful component (ICA000) is the noise from blinking, which is removed in all datasets.*

The distribution of EEG data during the squeeze period follows a slightly more Gaussian trend when compared to the distribution of EMG data, as the EEG datasets do not demonstrate significant skews in either direction and has a kurtosis closer to that of a Gaussian distribution. In addition, the range of variances in the EEG datasets (1.65e-11 to 2.25e-11) does not vary as widely as the variances in the EMG datasets, as discussed in section 3.3. Finally, as expected, each dataset has a near-zero mean.

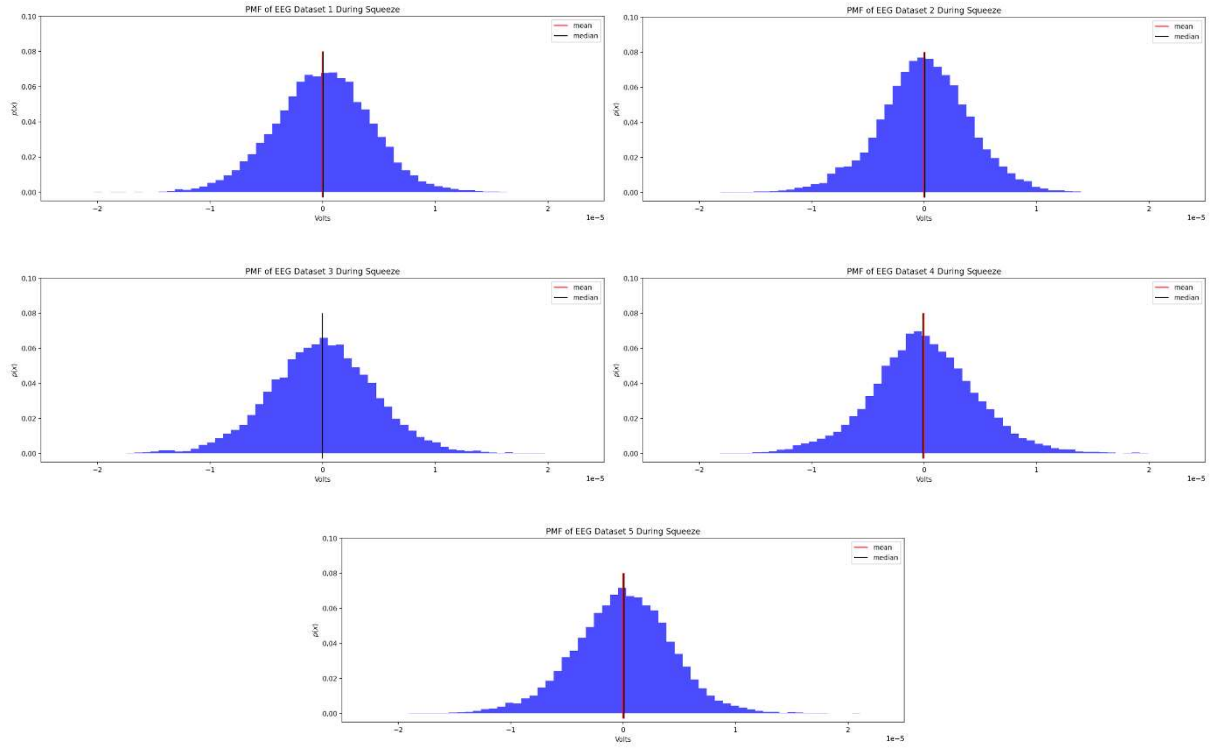


Figure 3-21*. Probability mass functions of EEG datasets, with means and medians marked in red and black, respectively.*

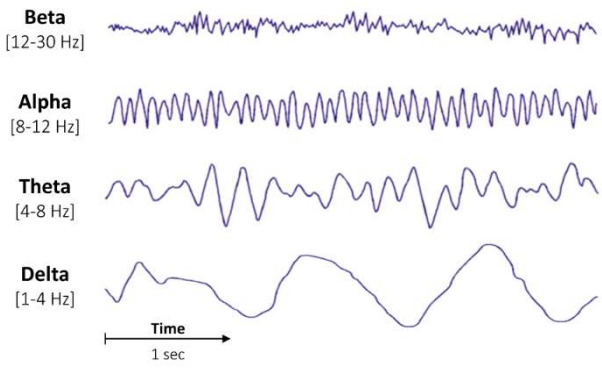


***Figure 3-22**. Visual representation of the four low-frequency EEG bands.*

While time-domain analyses provide some insight into the properties of the EEG data collected in this study, frequency-domain representations of EEG data can offer a more complete picture of brain activity.

From Figure 3-23, the built-in notch filter at 60 Hz is immediately noticeable. This filter suppresses frequencies between 58 and 62 Hz, lowering their amplitude compared to adjacent frequencies. Also noticeable is a concentration of energy in the lower frequencies, from 2 Hz to 30 Hz. This is a common characteristic of EEG signals - frequencies below 30 Hz typically carry the majority of the energy.

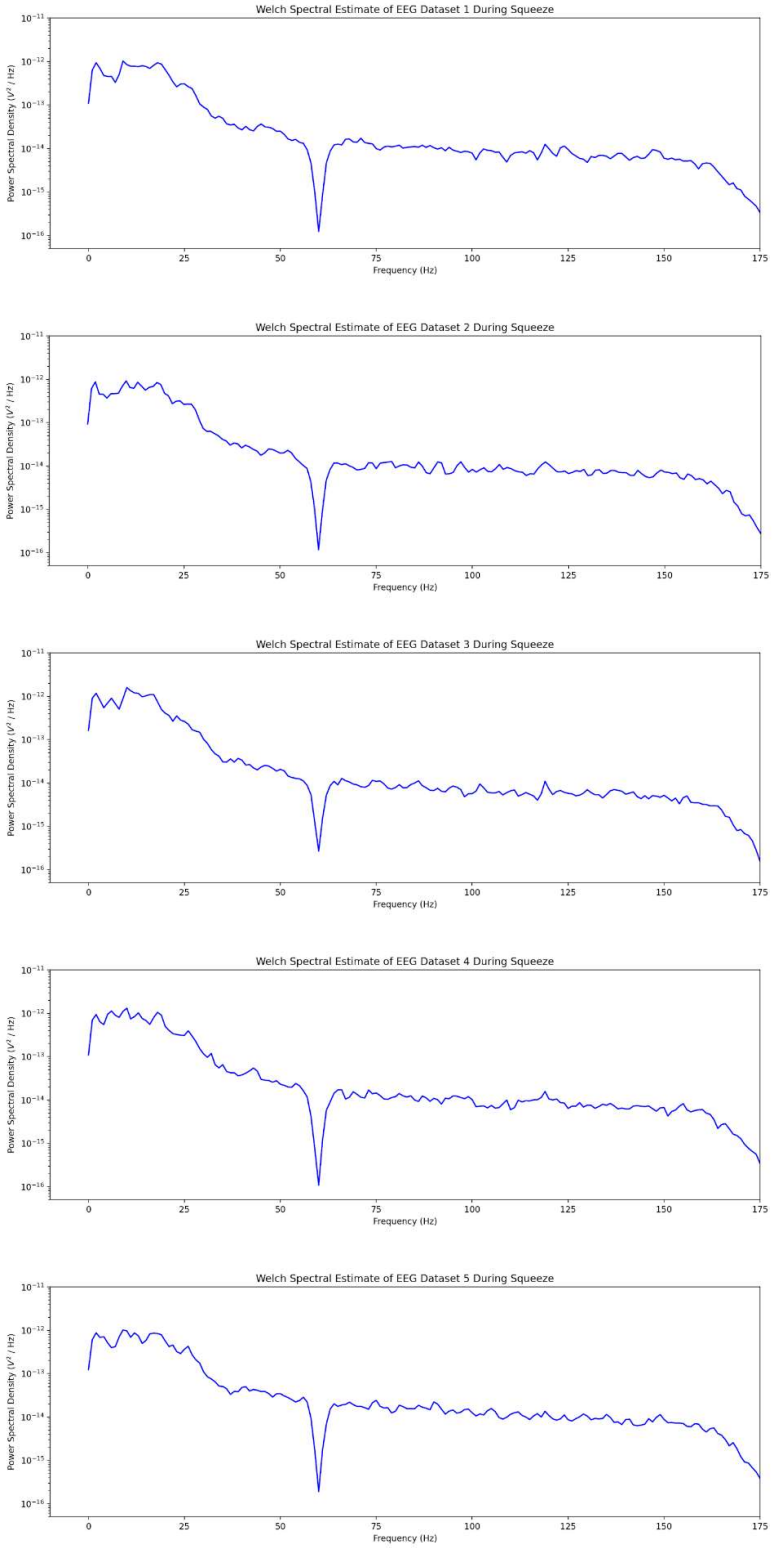


***Figure 3-23**. Power spectral density of processed EEG data.*

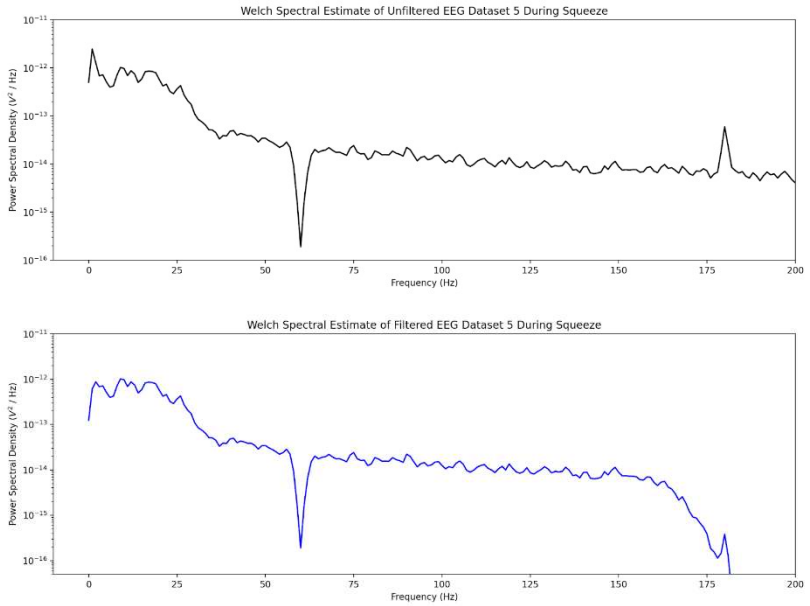


***Figure 3-24**. Depiction of changes in the frequency spectrum after preprocessing.*

Observing how the EEG data's power spectral density changes with movement tasks is crucial in obtaining a better understanding of how brain data relates to movement information. In Figure 3-25, spectrograms of each EEG dataset help illustrate this relationship. The squeeze region is indicated by the vertical black lines.

As with the power spectral density plots, it is immediately clear that there is much more power in the 2-30 Hz range than in higher frequencies. However, the spectrograms also show that this is consistent regardless of task. Occasional bursts of power in the higher frequencies can be seen at some time points, but these bursts do not seem to provide timing information regarding the squeeze period. However, closer inspection of the alpha band (EEG frequencies between 8-12 Hz) and low beta band (12-25 Hz) reveals that there is less power in that range of frequencies when the subject engages in the squeeze movement. During the non-squeeze period, especially after the squeeze, bursts of power can be observed near 10 Hz.

This finding was further verified by observing the results of singular value decomposition (SVD) performed on the EEG spectrogram matrix. SVD results in three matrices whose interpretations depend on the axes of the original matrix. In the case of the spectrogram matrix, the three resulting matrices correspond to the dominant 1) *frequency* patterns, 2) their respective *contribution* to the overall structure of the spectrogram (singular values), and 3) the prevalence of that frequency pattern in *time*. To quantify the importance of each spectral-temporal mode identified via SVD, the singular values were normalized by the sum of all singular values. This provides a measure of *relative* contribution, expressed as a percentage of the total.

Shown in Figure 3-26 are the most powerful frequency patterns in EEG datasets 1 and 2, with a relative magnitude of about 27.6% and 22%, respectively. The frequency pattern is mostly characterized by a strong dip near 10 Hz, which is in the center of the alpha band in EEG data. From the respective time components in the bottom of Figure 3-26, it can be seen that this dip is reversed into a burst in alpha band power during the non-squeeze period (generally outside of the 10-45 seconds time range) and is not strong during the squeeze. This is a consistent observation in all five datasets, which indicates a strong relationship between alpha-band power in EEG data collected at the C3 electrode and squeezing movements in the right hand. However, this cannot be considered a deterministic relationship - these bursts in power do not always line up perfectly with the squeeze/non-squeeze periods.

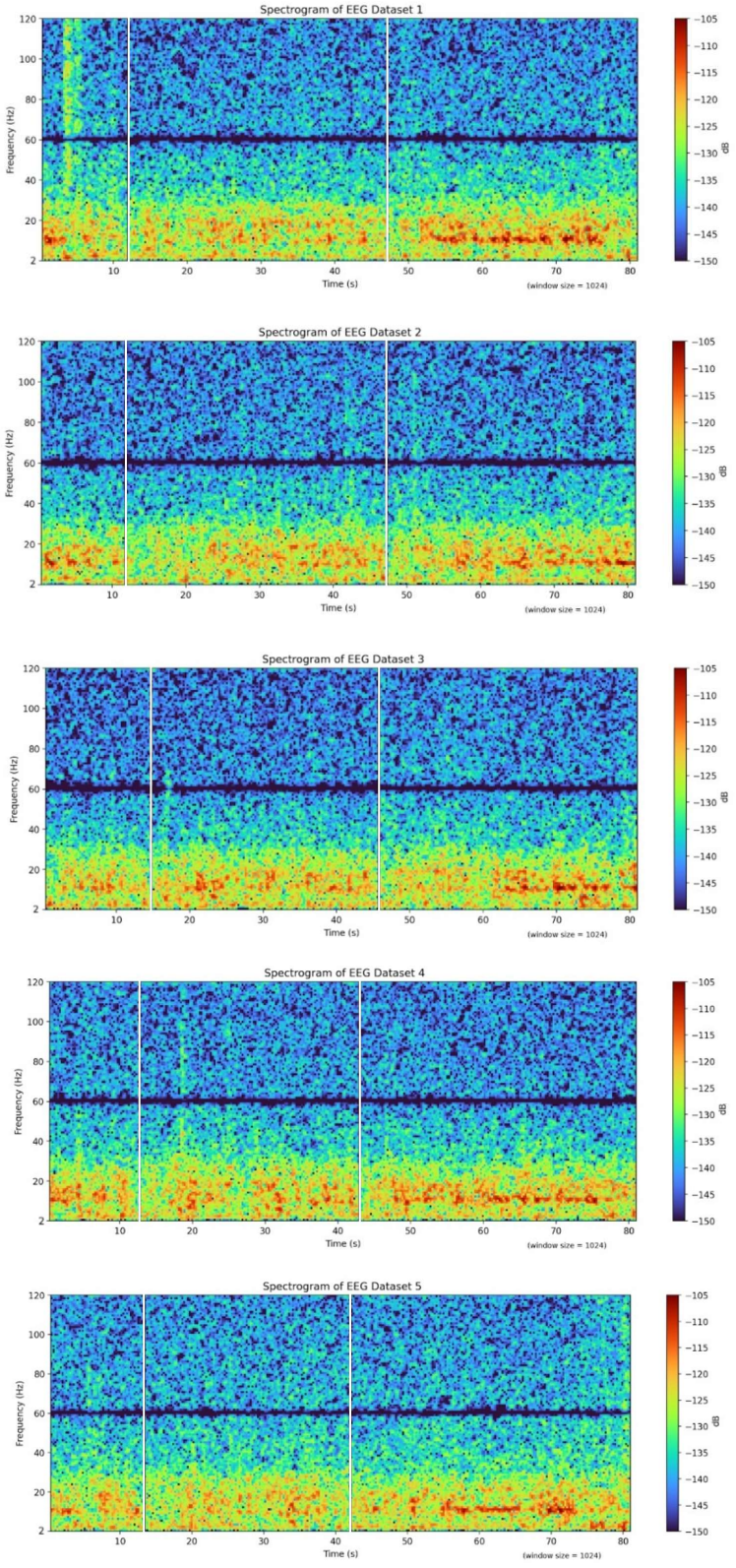


***Figure 3-25**. Spectrograms of EMG datasets using a 1024ms window and 75% overlap.*

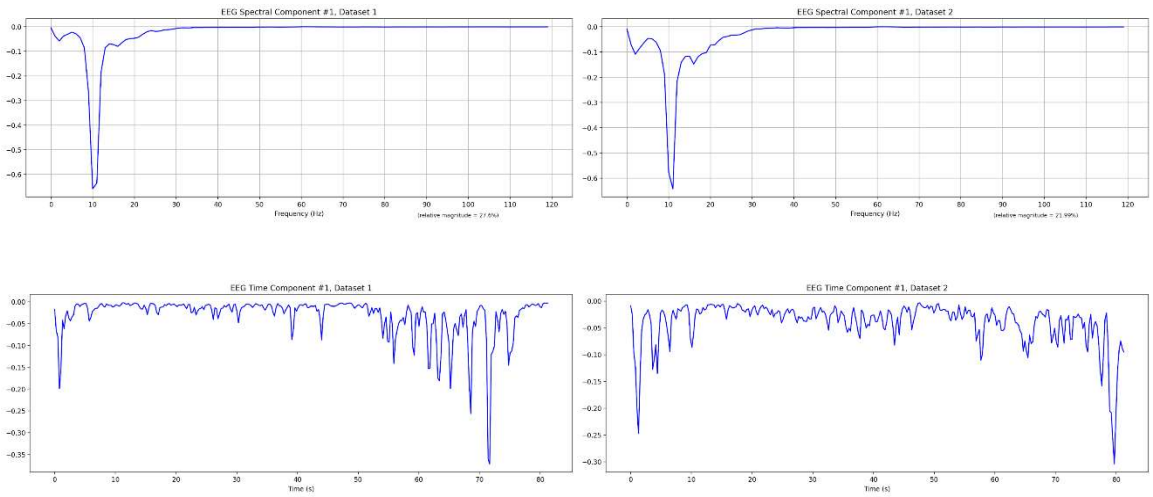


***Figure 3-26**. Most significant spectro-temporal components in EEG datasets 1 (left) and 2 (right).*

Finally, as EEG data is generally understood to be a correlated random process, it is proper to characterize the data by estimating the autocorrelation function for each dataset. As with the EMG data, the analysis is restricted to the squeeze region of the EEG data. The normalized autocorrelation estimates show that the autocorrelation of EEG data sharply decreases with lag and then rises to converge to zero, similar to the shape of the EMG autocorrelation function. However, the EEG autocorrelation function takes longer to converge to zero - the initial decrease in autocorrelation lasts until approximately 30 ms of lag, and in most cases, the autocorrelation doesn't return to zero until about a 40 ms lag. This is partly due to the stronger prevalence of low frequencies in EEG data, especially when compared to the more uniform distribution of power across frequencies in EMG data. In addition, an interesting synchronization of the autocorrelation functions of all five EEG datasets is observable between lags of 115 and 125 ms. This aligns with the periodicity of alpha-band oscillations (8–12 Hz), which were found to be related to movement tasks. The recurring autocorrelation structure at this specific lag may reflect the association between alpha-band activity and motor processes.

Finally, the comparison of the normalized autocorrelation function estimates shows that the autocorrelation of squeeze-related EEG data is less consistent across EEG datasets than the autocorrelation functions of the EMG datasets (as presented in Figure 3-12), indicating slightly more variability in the stochastic properties of the EEG data used in this research.

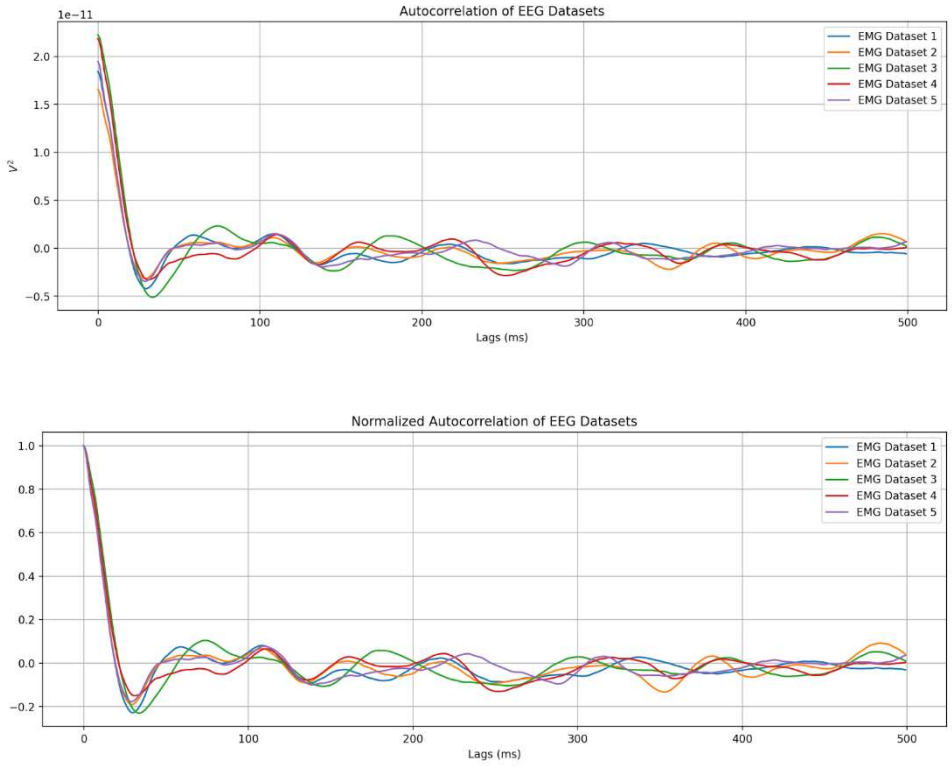


Figure 3-27*. Top) Autocorrelation estimates of EEG datasets. Bottom) Plot of normalized autocorrelation for easier comparison of autocorrelation shape.*

4. EEG and EMG Models

4.1 Linear Predictive Modeling Methods

In this chapter, we investigate the utility of linear predictive models in characterizing EMG and EEG data during prescribed movement. Linear predictive models have been widely used in fields like speech processing for their ability to compress the information contained in stationary random processes like short speech sounds. LP models work especially well in sounds (and other types of data) that can be modeled as autoregressive random processes - processes that are non-deterministic functions of their previous samples.

Autoregressive linear predictive models can also be understood as systems that filter Gaussian white noise into a given random process with desired statistical, autocorrelation, and power spectral density characteristics. In this understanding, the filters that define this system are all-pole filters.

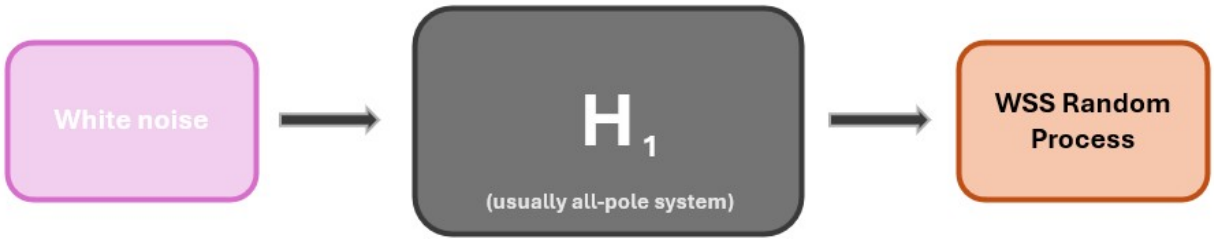


***Figure 4-1**. White-noise-excitation representation of linear predictive modeling.*

The effectiveness of all-pole linear predictive models often depends on the degree to which the source of the random process exhibits autoregressive behavior, as well as the consistency of the source's autoregressive properties. In literature, EEG and EMG data have been modeled as autoregressive random processes, under certain circumstances. As mentioned in the background section of this thesis, studies by Anjum et al. (2020) and Carotti et al. (2007) established all-pole models for EEG and EMG data under specific conditions, successfully characterizing electrophysiological data with meaningful features.

The calculation of filter coefficients for an all-pole model involves a set of linear equations that relate previous autocorrelation values to an autocorrelation value at a future lag. These equations are referred to as the Yule-Walker equations, as shown in Figure 4-2 (Therrien, 1992). Because the exact autocorrelation function of a given random process can only be *estimated* based on collected data, the exact calculation of the autocorrelation matrix used in the Yule-Walker equations varies from method to method. In the Autocorrelation Method of all-pole modeling, the autocorrelation matrix is estimated as shown in Figure 4-3.

$$\underset{(P+1)\,x\,(P+1)}{\begin{bmatrix} R_x(0) & R_x(1) & R_x(2) & \dots & R_x(P) \\ R_x(-1) & R_x(0) & R_x(1) & \dots & R_x(P-1) \\ R_x(-2) & R_x(-1) & R_x(0) & \dots & R_x(P-2) \\ \vdots & \vdots & \vdots & & \vdots \\ R_x(-P) & R_x(-P+1) & R_x(-P+2) & \dots & R_x(0) \end{bmatrix}} \cdot \underset{(P+1)\,x\,1}{\begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix}} = \underset{(P+1)\,x\,1}{\begin{bmatrix} {\sigma_\omega}^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}$$

Figure 4-2*. Yule-Walker equations.*

$$X = \begin{bmatrix} 0 & 0 & \dots & 0 & X(0) \\ 0 & 0 & \dots & X(0) & X(1) \\ \vdots & \vdots & & \vdots & \vdots \\ X(0) & X(1) & \dots & X(P-1) & X(P) \\ X(1) & X(2) & \dots & X(P) & X(P+1) \\ \vdots & \vdots & & \vdots & \vdots \\ X(N_s-P-1) & X(N_s-P) & \dots & X(N_s-2) & X(N_s-1) \\ X(N_s-P) & X(N_s-P+1) & \dots & X(N_s-1) & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ X(N_s-1) & 0 & \dots & \dots & 0 \end{bmatrix}_{N_s \text{ x } (P+1)}$$

$$R_x = X^T X$$

Figure 4-3. *The calculation of the autocorrelation matrix used in the autocorrelation method.*

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_P \end{bmatrix}_{\text{P x 1}} = \begin{bmatrix} R_x(0) & R_x(1) & \dots & R_x(P-1) \\ R_x(1) & R_x(0) & \dots & R_x(P-2) \\ \vdots & \vdots & & \vdots \\ R_x(P-1) & R_x(P-2) & \dots & R_x(0) \end{bmatrix}^{-1}_{\text{P x P}} \cdot \begin{bmatrix} -R_x(1) \\ -R_x(2) \\ -R_x(3) \\ \vdots \\ -R_x(P) \end{bmatrix}_{\text{P x 1}}$$

$$\sigma_\omega{}^2 = \underset{\text{1 x (P+1)}}{[R_x(0) \quad R_x(1) \quad R_x(2) \quad \dots \quad R_x(P)]} \cdot \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_P \end{bmatrix}_{\text{(P+1) x 1}} = b_0{}^2$$

Figure 4-4. *Direct equations used in the calculation of filter coefficients (top) and white noise variance/scaling factor (bottom), derived from the Yule-Walker equations.*

Once the autocorrelation matrix has been estimated, the Yule-Walker equations can be manipulated to solve for the filter coefficients and the white noise variance, as seen in Figure 4-4. The white noise variance term, which can be represented by σ_w or b_0, can be understood as a scaling factor that scales the power of the input white noise (with unit variance) into a power

level similar to that of the signal being modeled. As illustrated in Figure 4-5, this term can be understood as the numerator in the z-domain representation of the filter's transfer function.

The Autocorrelation Method has a benefit over other AR model estimation techniques in that it guarantees stability and causality in the resulting all-pole filter. This is due to the positive semidefinite and Toeplitz properties of the autocorrelation matrix used to solve for the filter coefficients. However, a tradeoff of this method is that it is a deterministic method - the resulting model depends heavily on the specific data segment (from Figure 4-3) used to calculate the autocorrelation matrix. This sensitivity can be mitigated by using a sufficiently long data segment, which improves the representativeness of the estimated autocorrelation matrix. Because the duration of the squeeze period varies across EEG/EMG datasets (ranging from 28 seconds to 37 seconds, as described in section 3.3), the first 25 seconds of each squeeze segment were used to construct the EEG and EMG models discussed in the following subsections. This standardized length prevents possible confounding factors that may come from using different amounts of data in fitting the all-pole EEG/EMG models.

4.2 EMG Models

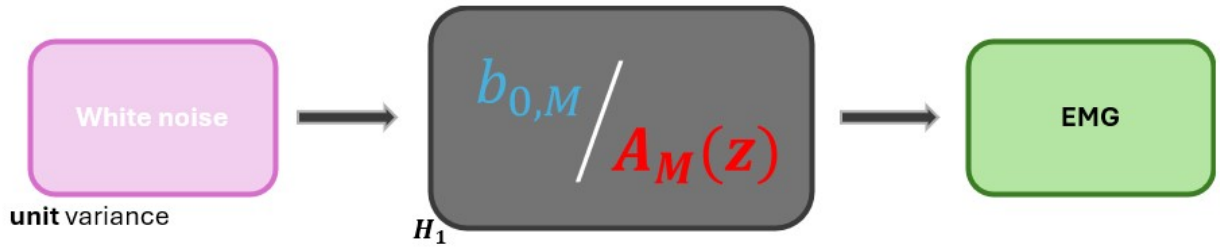


Figure 4-5. White-noise-excitation representation of EMG AR-modeling.

The task of signal modeling can often be reframed as shaping the frequency domain of white noise into that of a desired signal. Similarly, the accuracy of a signal model can be understood as the frequency domain difference between the model estimate and the desired signal. A standard method of comparing the frequency domain difference between the model estimate and the desired signal is the spectral flatness method, which inverts the white-noise-excitation representation to evaluate the "whiteness" of the inverse filter's output, given the original data as input (Figure 4-6). This can be understood as a "white noise estimate" – a more accurate model will closely approximate white noise.

The defining frequency domain characteristic of white noise is its "flatness" across frequencies, so the spectral flatness equation included in Figure 4-7 is used to assess the uniformity of the inverse filter's output across the frequency domain. For example, if high power is only observed in a specific, narrow frequency range of the resulting white noise estimate, the spectral flatness will be close to zero. If the distribution of power in the frequency domain is quite uniform, then the spectral flatness of the white noise estimate will be close to 1. Ideally, the inverse of a successful model will return white noise with a spectral flatness close to 1.

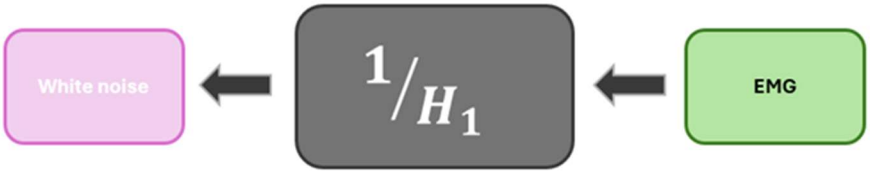


***Figure 4-6**. Depiction of the inverse filter used to calculate spectral flatness.*

$$Spectral\ Flatness = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\left(\frac{\sum_{n=0}^{N-1} x(n)}{N}\right)} = \frac{geometric\ mean}{arithmetic\ mean}$$

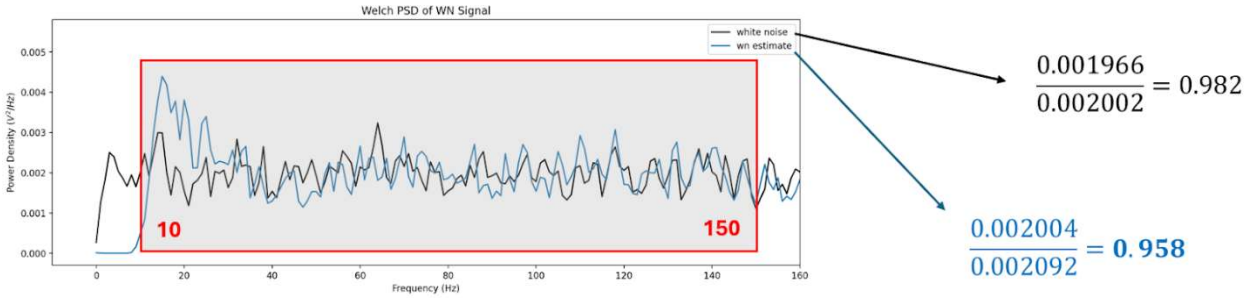


***Figure 4-7.** Illustration of the spectral flatness calculation.*

Order selection for the EMG model was completed by evaluating the spectral flatness of the model as the model order increased, as shown in Figure 4-8. As expected, performance increases with the number of parameters used. However, increases in performance seem to plateau past a model order of 11 or 12. This indicates that additional parameters do not serve to improve the model estimate significantly, which marks the characteristic of an optimal model order as described in (Makhoul, 1975). The model converges to a strong spectral flatness of 0.95.

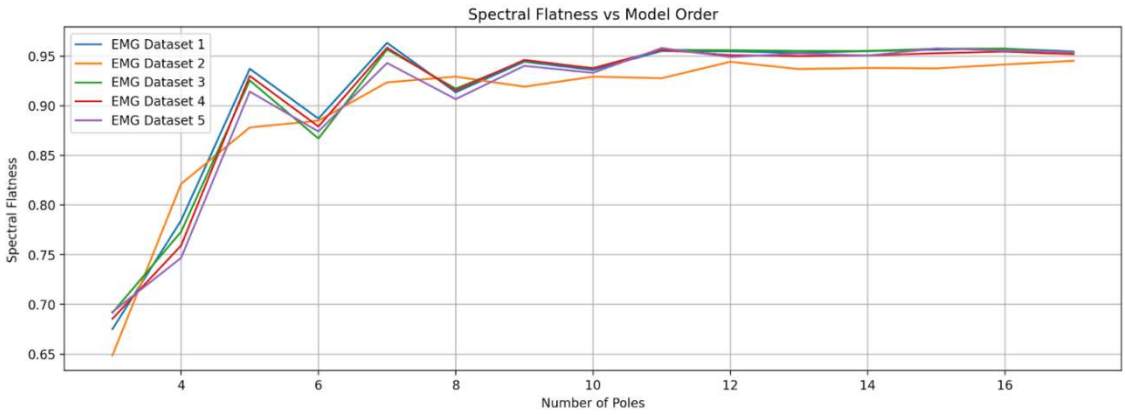


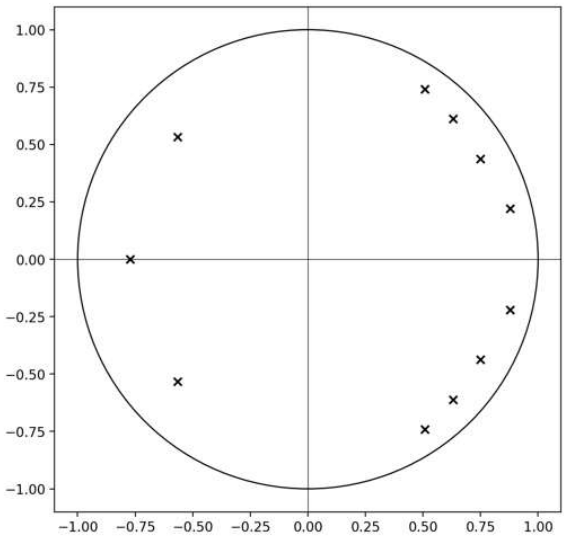
***Figure 4-8**. Spectral flatness as it depends on the order of the AR model. Ideally, the performance of the model converges past a certain model order.*

With the exception of dataset 2, the spectral flatness vs model order curve in Figure 4-8 has an interesting property in that performance seems to be better for models with an odd number of poles, relative to adjacent orders that have an even number of poles. One possible explanation for this phenomenon is that odd-order models allow for the inclusion of a real-valued pole alongside pairs of complex-conjugate poles, whereas even-order models may lack this flexibility, as all poles must be arranged in complex-conjugate pairs. This suggests that including a dominant real pole (with no resonant frequency) in the EMG model may help provide a more appropriate, smoother representation of the EMG signal structure. However, this phenomenon is observed to a much less significant degree in EMG dataset 2, so this property may not be equally prominent in all such EMG activity.

The 11-pole AR model was chosen as the optimal EMG model for the given datasets because additional parameters result in severely diminishing returns past a model order of 11 poles. Figure 4-9 illustrates the pole-zero plot and difference equation that define the EMG model fitted using EMG dataset 5. Clearly this model is stable, as all poles are within the unit

circle. The impulse response and step response in Figure 4-10 confirm this property of the model. In addition, the poles in the right side of the unit circle are closer to the edge of the unit circle than the poles in the left side of the unit circle. This suggests a focus on power in the low-frequency content of EMG data, which is expected.

Additionally, the impulse response of the EMG model — which is designed to shape white noise into an EMG-like signal — closely resembles the empirical autocorrelation of squeeze-related EMG signals presented in Section 3.3. This is an expected result, as the autocorrelation of the model output is the convolution of the filter's impulse response with the autocorrelation of the input white noise, which is a Dirac delta function centered at n = 0.



$$\begin{aligned}\mathrm{EMG}[n] = (5.81 \cdot 10^{-7}) \cdot \mathbf{WN}[n] &+ 3.63\,\mathrm{EMG}[n-1] \\ &- 5.42\,\mathrm{EMG}[n-2] + 3.37\,\mathrm{EMG}[n-3] \\ &+ 0.50\,\mathrm{EMG}[n-4] - 1.59\,\mathrm{EMG}[n-5] \\ &- 0.16\,\mathrm{EMG}[n-6] + 1.03\,\mathrm{EMG}[n-7] \\ &+ 0.02\,\mathrm{EMG}[n-8] - 0.93\,\mathrm{EMG}[n-9] \\ &+ 0.69\,\mathrm{EMG}[n-10] - 0.18\,\mathrm{EMG}[n-11]\end{aligned}$$

***Figure 4-9**. Pole-zero plot (left) and difference equation (right) of the 11-pole EMG model.*

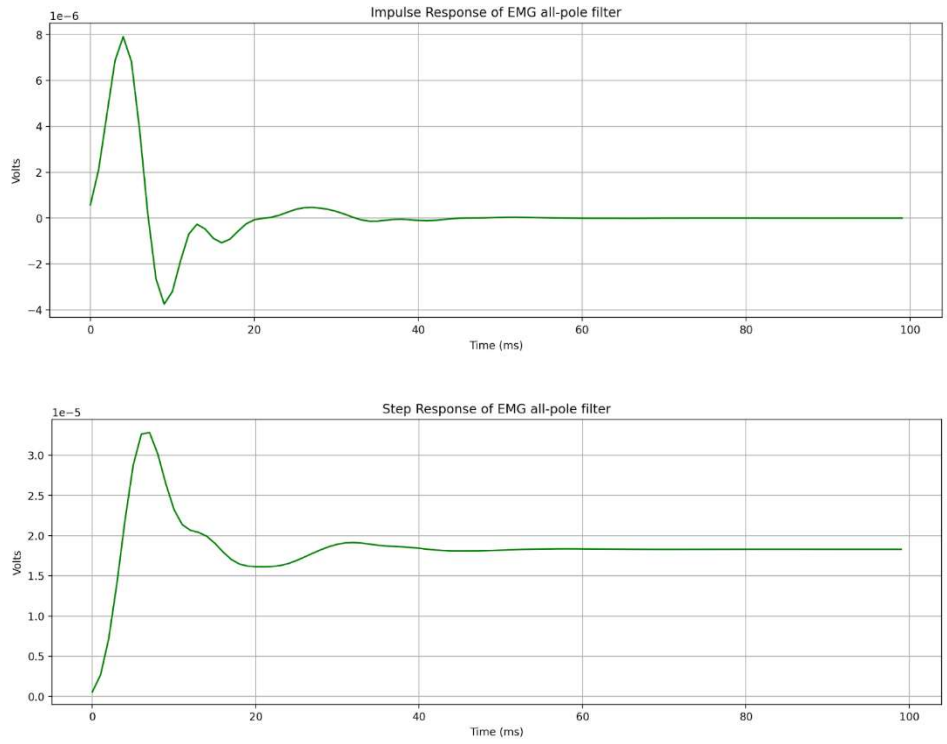


***Figure 4-10**. Impulse response (top) and step response (bottom) of the 11-pole EMG model.*

After the EMG model was calculated and stability was ensured, the EMG estimate was generated by feeding unit-variance white noise into the all-pole filter. The autocorrelation of the EMG estimate produced by the 11-pole EEG model, shown in the top of Figure 4-11, closely matches that of the real EMG squeeze data up until approximately 14 ms of lag. It is expected that an nth-order AR signal model will be able to replicate the desired signal's autocorrelation function up to an nth lag, but this model shows an added ability to match the autocorrelation function closely, even past this point. Additionally, the power of the EMG estimate signal is nearly identical to that of the real EMG squeeze data, as evidenced by the matching autocorrelation functions at lag 0 ms. It can also be seen that the autocorrelation of the EMG estimate converges to zero with increasing lag, as expected.

Between 15 ms and 25 ms of lag, however, the EMG estimate's autocorrelation function slightly overshoots the autocorrelation of the real data. Although this contributes to the model error, it can be seen from Figure 3-12 that the autocorrelations of the EMG datasets converge to zero in a variety of ways past a lag of 20 ms. This suggests that model orders should not exceed 20 parameters , as attempts to fit the autocorrelation of the EMG estimate past 20 ms of lag may lead to dataset-specific overfitting (due to the 1 kHz sampling rate, each lag term in the difference equation corresponds to 1 ms of lag).

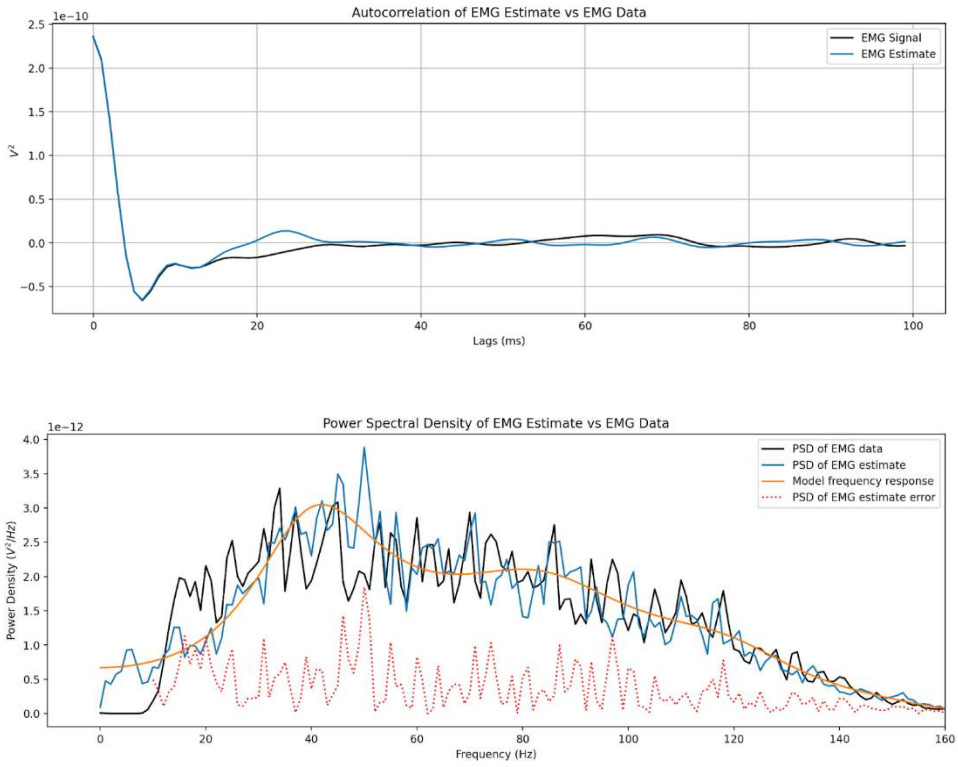


***Figure 4-11**. Comparison of the autocorrelation function (top) and power spectral density (bottom) in the EMG model estimate and real EMG data from dataset 5.*

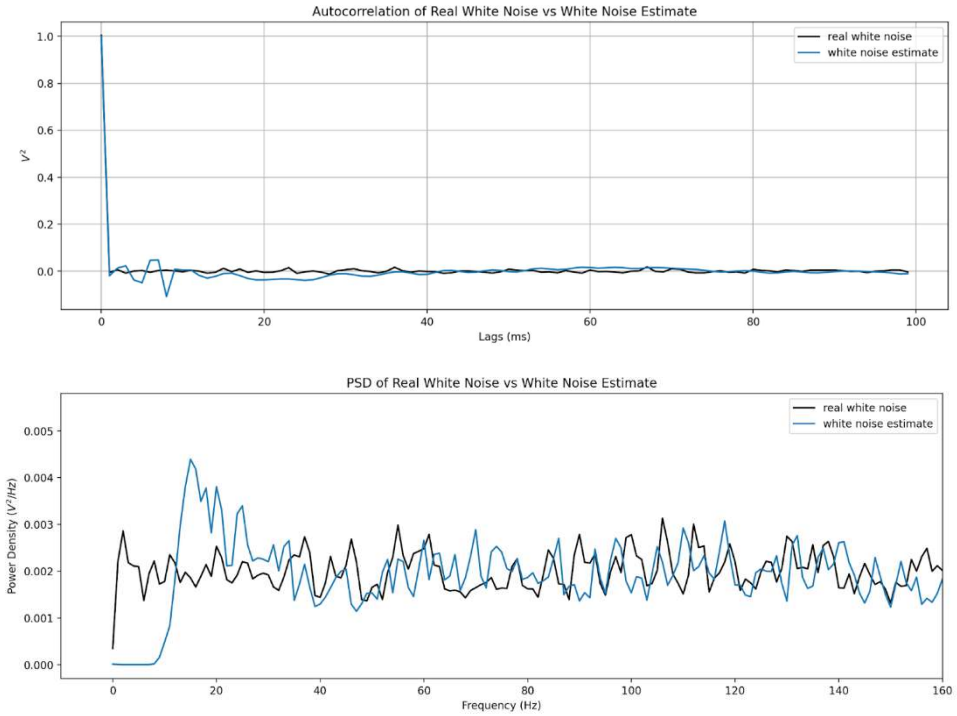


***Figure 4-12**. Comparison of the autocorrelation function (top) and power spectral density (bottom) in the EMG inverse model's white noise estimate and real white noise.*

In the frequency domain, barring inherently random fluctuations in the exact power density at each frequency, the overall shape of the EMG estimate's power spectral density closely matches that of the real EMG data. The trend of smooth decreases in power with increases in frequency is replicated in the frequency distribution of the EMG estimate.

Errors in the 10-30 Hz range can be traced back to the EMG model's attempt to replicate the steep built-in highpass filter at 10 Hz in the real EMG data. Because of this unavoidable property of the EMG data, the EMG estimate consistently undershoots the power in the 10-30 Hz range, as it tries to replicate both the attenuated power in the lowest frequencies and the high power in the adjacent frequencies. Otherwise, the error in the EMG estimate's power spectral density, calculated by taking the absolute value of the difference between the power spectral

density of the estimate and the power spectral density of the real EMG data, is largely random and independent of frequency.

As mentioned earlier, signal models can be evaluated by examining the output of their inverse. When the EMG data is fed into the inverse of the calculated EMG model, the output has a frequency distribution very similar to white noise, with a spectral flatness of approximately 0.96. As seen in Figure 4-12, the main source of errors in the inverse model's white noise can be found in the 10-30 Hz range, which corresponds to the subpar match between the EMG estimate and the real EMG data in the same frequency range. Despite this, the spectral flatness of the white noise produced by the inverse EMG model closely follows that of real white noise, which typically has a spectral flatness that ranges from 0.98 to 1.

Although the model replicates spectral and temporal properties of the EMG data quite well, a possible limitation of this AR model can be found when comparing the distribution of the EMG estimate to the distribution of real EMG data. The slightly left-skewed distribution of EMG data exhibits a kurtosis of 5.1, which indicates a non-Gaussian, leptokurtic distribution. In contrast, the EMG estimate exhibits a near-Gaussian distribution with a kurtosis of 3 and very little skew. This discrepancy is largely due to the standard use of Gaussian-distributed white noise as an input to the all-pole filter. Future work could involve the use of a carefully skewed or heavy-tailed distribution to better match the non-Gaussian characteristics of real EMG data.

In order to quantify the dissimilarity between these two distributions, the Kolmogorov-Smirnov (KS) test can be used to assess the likelihood that these two datasets come from the same distribution. The null hypothesis of the KS test is that the two distributions come from the same source. The KS test that compares the EMG estimate and the EMG dataset results in a very low p-value of 1e-41, which confirms the initial qualitative observation of a less-than-ideal

match in distribution. Otherwise, the near-zero means and standard deviations of the EMG estimate (mean = -9.8e-8, standard deviation = 1.51e-5) and real EMG data (mean = 3.5e-10, standard deviation = 1.54e-5) are very similar.

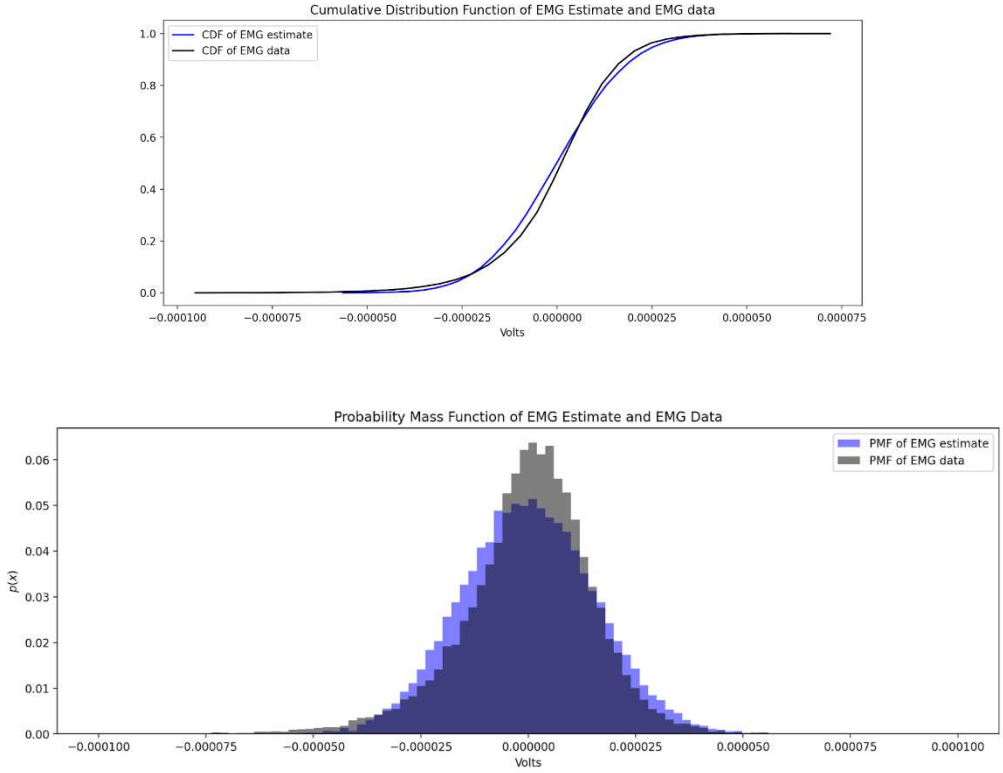


Figure 4-13. *Cumulative distribution function and probability mass function of the EMG estimate, as compared to that of EMG dataset 5.*

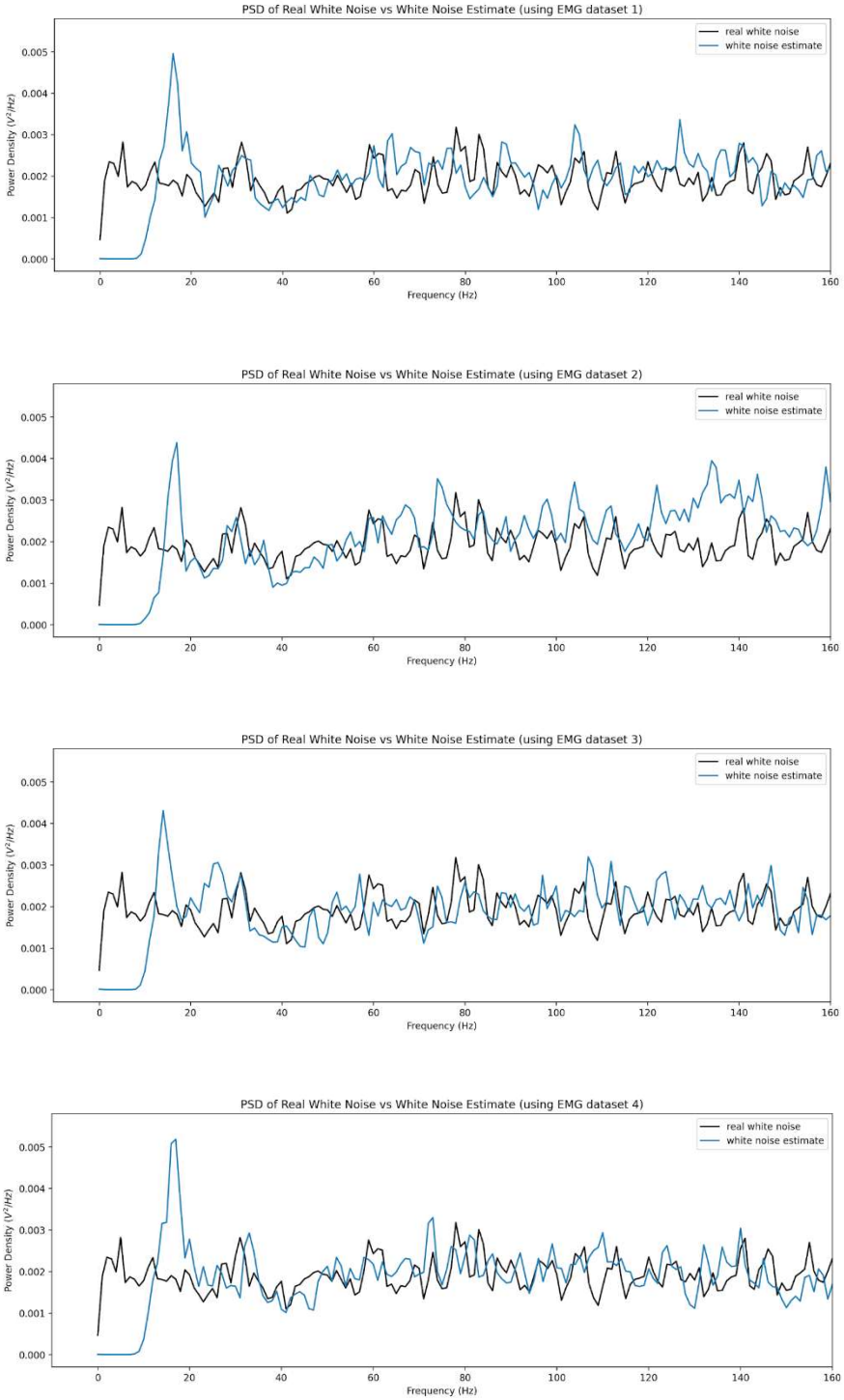


***Figure 4-14**. Power spectral densities of white noise resulting from feeding each EMG dataset into the EMG model built using EMG dataset 5.*

While the model demonstrates strong performance when applied to the dataset it was trained on (dataset 5), it is important to evaluate its generalizability across the remaining EMG datasets. Since the model is inherently tailored to reproduce the characteristics of dataset 5, its effectiveness on other datasets is closely tied to the similarity of the datasets. As observed in section 3.3, the EMG datasets share comparable normalized autocorrelation and power spectral density shapes, but they differ in overall signal power. In order to fairly evaluate model performance across these datasets and mitigate the influence of scale mismatches, all datasets were normalized to have the same overall power as EMG dataset 5. Using spectral flatness as a performance metric, the model achieved values of 0.961, 0.929, 0.962, and 0.957 when applied to the other four normalized datasets, respectively. These results indicate that the constructed EMG model can generalize enough to model the EMG signal across each of the movement-related EMG datasets, and does so with sufficient performance.

4.3 EMG Models

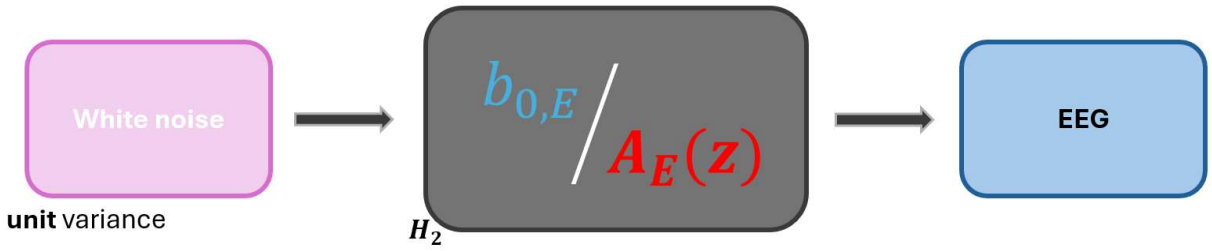


Figure 4-15. White-noise-excitation representation of EEG AR-modeling.

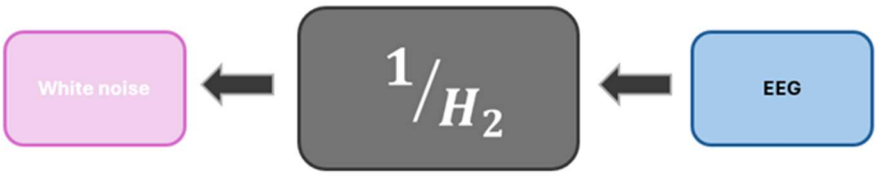


Figure 4-16. Depiction of the inverse EEG filter used to calculate spectral flatness for each model.

As with the previously discussed EMG model, the spectral flatness method can be used to determine the optimal signal model for EEG data. As seen in the spectral flatness vs model order curve in Figure 4-17, EEG model performance has much more variance across EEG datasets, especially when compared to the same curve for the EMG models (Figure 4-8). Additionally, model performance does not consistently depend on whether there are an even/odd number of poles - some datasets are better approximated with an even number of poles, like EEG dataset 4, some are better approximated with an odd number of poles, like EEG dataset 1, and others have no clear dependence on this even-odd distinction. This is likely due to the high variability and

intricacy in the frequency domains across EEG datasets, especially as compared to the relatively consistent and simple frequency domain shape observed in EMG data.

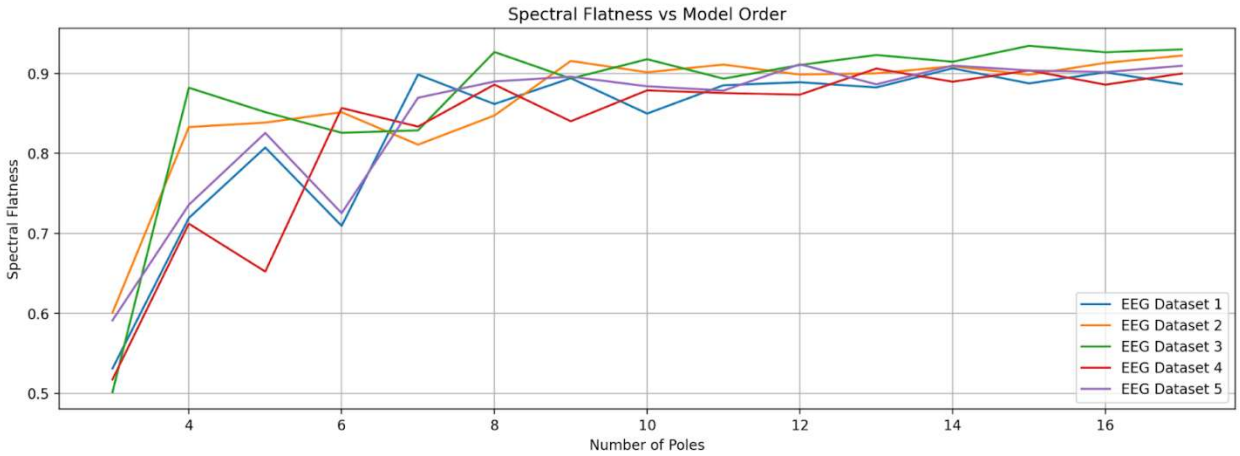


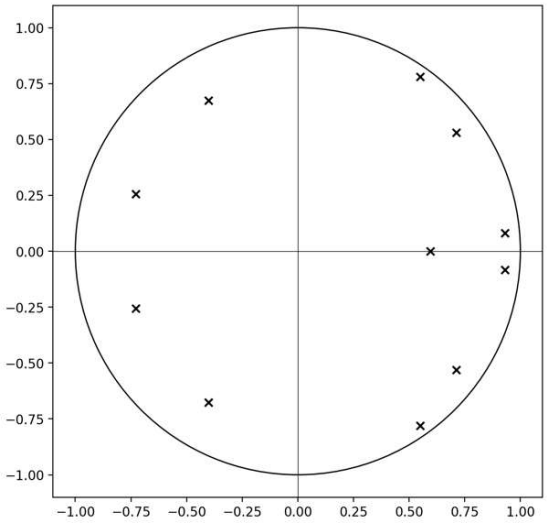
***Figure 4-17**. Spectral flatness of the EEG models as model order increases.*

In addition, Figure 4-17 also reveals that EEG model performance seems to converge near an average spectral flatness of 0.9, as opposed to 0.95 in the EMG models. This can be related to the relative difficulty of describing the more complex peaks in EEG data using AR models, especially within the narrow bandwidth of 2-30 Hz, which contains the bulk of the EEG signal power.

Additional parameters result in diminished increases in performance around a model order of 8 poles. However, the wide variability of the performance at this model complexity across EEG datasets, which has a range of 0.07, indicates that this model order is not equally optimal for all datasets. At a model order of 11, the range of performance narrows to approximately 0.03, and this range stays consistent in all following model orders. Therefore, an

all-pole model with 11 poles was also chosen as the optimal model for EEG data during the voluntary squeeze motion. However, it is important to note that model performance continues to increase slightly, even past 11 poles. This suggests that, unlike the EMG model, the optimal AR model for EEG data may not fully capture the dynamics of EEG data. Significant improvements beyond a spectral flatness of 0.9 may require additional terms - possibly non-AR terms - in future iterations of the EEG signal model.

The pole-zero plot (Figure 4-18) of the optimal all-pole EEG model indicates stability and causality in the filter representation of the model, and this filter stability is confirmed with the convergence of the impulse response and step response, as shown in Figure 4-19. The proximity of the right-side poles to the perimeter of the unit circle indicates a much stronger focus on low-frequency power than on high-frequency power, which is expected based on the distinct properties of the power spectral density of EEG data. The $b_{0,E}$ scaling factor in the difference equation of the EEG signal model is less than that of the EMG signal model ($b_{0,M}$), which correctly indicates that EEG data is of lower power than EMG data.



$$\begin{aligned}\text{EEG}[n] = (2.12 \cdot 10^{-7}) \cdot \mathbf{WN}[n] &+ 2.71\ \text{EEG}[n-1] \\ &- 2.55\ \text{EEG}[n-2] + 0.32\ \text{EEG}[n-3] \\ &+ 0.87\ \text{EEG}[n-4] + 0.03\ \text{EEG}[n-5] \\ &- 0.53\ \text{EEG}[n-6] - 0.18\ \text{EEG}[n-7] \\ &+ 0.36\ \text{EEG}[n-8] + 0.23\ \text{EEG}[n-9] \\ &- 0.41\ \text{EEG}[n-10] + 0.13\ \text{EEG}[n-11]\end{aligned}$$

***Figure 4-18**. Pole-zero plot (left) and difference equation (right) of the 11-pole EEG model.*

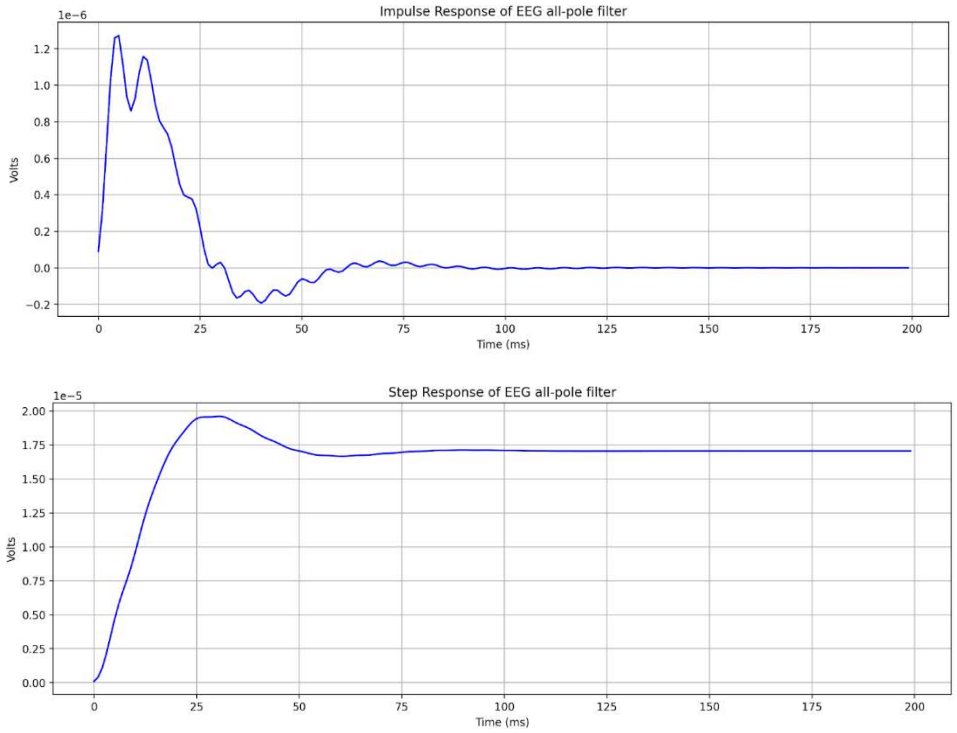


Figure 4-19. *Impulse response (top) and step response (bottom) of the 11-pole EEG model.*

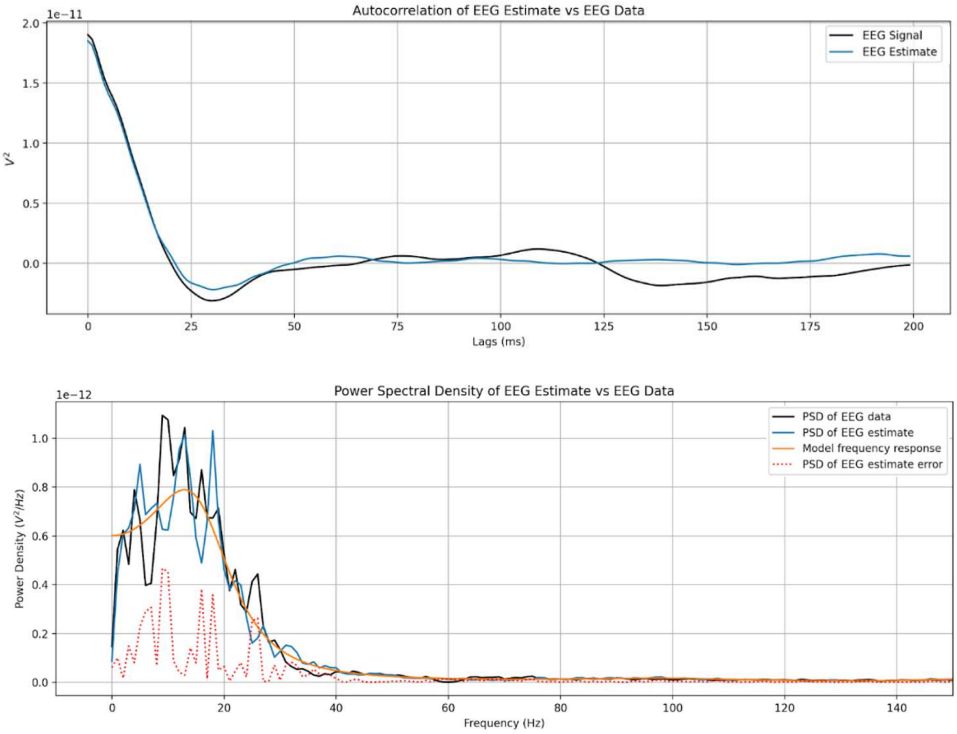


Figure 4-20. *Comparison of the autocorrelation function (top) and power spectral density (bottom) in the EMG model estimate and real EEG data from dataset 5.*

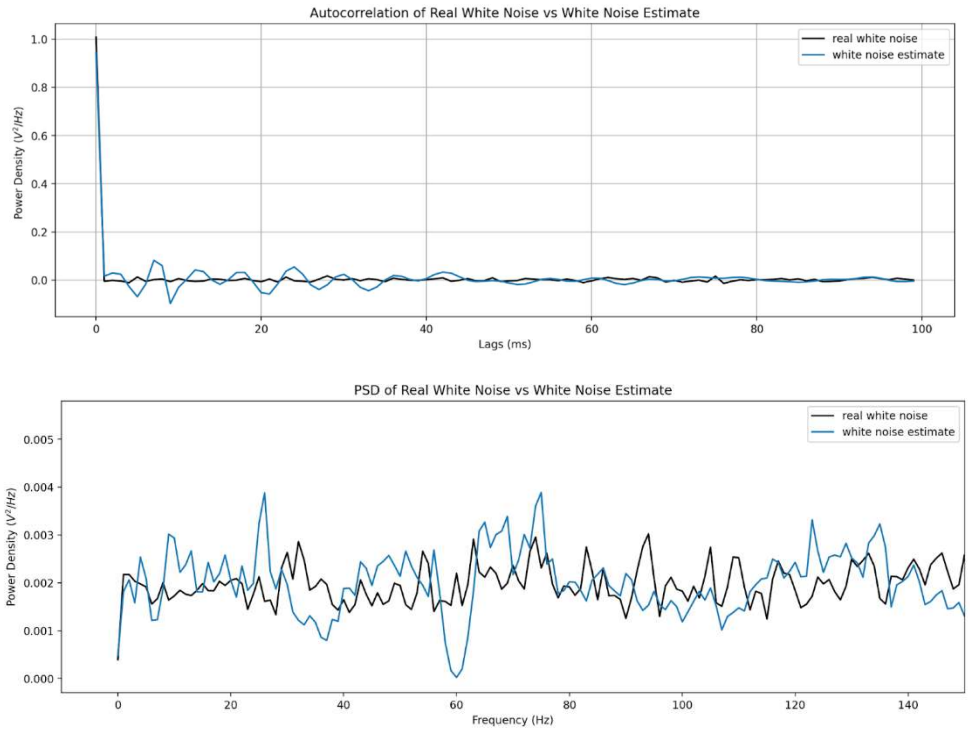


***Figure 4-21**. Comparison of the autocorrelation function (top) and power spectral density (bottom) in the EEG inverse model's white noise estimate and real white noise.*

The comparison of the autocorrelation function of the EEG estimate and the autocorrelation function of EEG dataset 5 confirms this conclusion, as the power of the EEG estimate (observed at lag 0 ms) is nearly equal to the power of the real EEG data (Figure 4-20). The two autocorrelation functions closely match up until 15 ms of lag. After 15 ms of lag, the autocorrelation of the estimate is not precisely identical to the real EEG data's autocorrelation function, but it follows the general trend up until 75 ms lag. This is acceptable because the normalized autocorrelations of the EEG datasets, as illustrated in Figure 3-27, demonstrate that the autocorrelation function in this range of lag is quite variable across EEG datasets. Therefore, there is no reason to further fit the model in this region. After 75 ms of lag, both autocorrelation functions oscillate around zero, with the estimate's autocorrelation function exhibiting slightly less deviation from zero than that of the real EEG data.

As mentioned, the distribution of power in the frequency domain of EEG data contains more pronounced peaks than in EMG data. Small, concentrated dips or peaks in power present the main challenge in using AR models to model EEG data. Increasing the order of the model to account for these intricate peaks did not succeed, as higher order EEG models also struggle to capture these sharp peaks and simultaneously maintain the overall shape of the EEG power spectral density. Although the dip and peak between 5 and 10 Hz were not captured perfectly in the EEG estimate, the general envelope of the EEG estimate's power spectral density closely follows that of the collected EEG data throughout the rest of the relevant frequency range.

Despite these struggles with sharp peaks in power, the inverse EEG model outputs white noise with a spectral flatness of approximately 0.902 for EEG dataset 5. While not quite as white as the output of the inverse EMG model, these results are largely skewed by the severe dip in power of the white noise estimate between the ranges of 57 Hz and 63 Hz. This effect is directly related to the 60 Hz notch filter built into the EEG data collection process. Because the EEG data contains little to no power in this bandwidth, the inverse filter also cannot produce any power at this bandwidth, resulting in a lack of control in this frequency region. In fact, a calculation of the spectral flatness of this white noise output in all the regions outside of the affected 57-63 Hz region (but still within the relevant bandwidth of 2-150 Hz) results in an improved spectral flatness of 0.945, confirming this observation.

Aside from this discrepancy, another difference between the power spectral density of real white noise and that of the inverse filter output is a small, sharp peak near 25 Hz, which is related to the EEG estimate's inability to model the extremely sharp peak in the real EEG data at that same frequency. Otherwise, the power spectral density of the inverse model output is uniformly distributed with random fluctuations, as is expected in practical white noise. The

autocorrelation function of the EEG estimate echoes this finding, as the slight coloration of the white noise estimate results in an autocorrelation function that quickly - but not immediately - converges to zero.

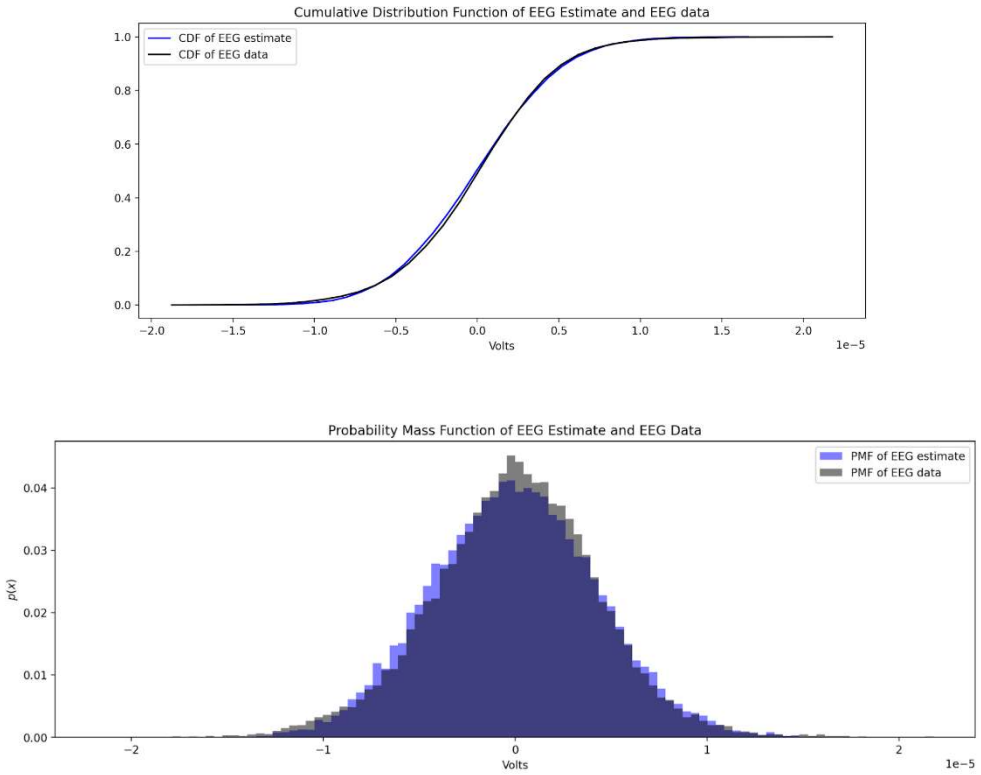


***Figure 4-22**. Cumulative distribution function and probability mass function of the EMG estimate, as compared to that of EMG dataset 5.*

Although the EEG model's performance in the frequency domain leaves some room for improvement, Figure 4-22 reveals that the distribution of the EEG estimate is similar to the distribution of real EEG data. This is mainly due to the near-Gaussian distribution found in EEG data and the Gaussian distribution in the EEG estimate, which is caused by the Gaussian white noise input.

Again, the KS test can be used to quantitatively compare the two distributions. Despite close visual alignment between the distribution of the model estimate and that of the real EEG data, the null hypothesis of the KS test (that the two distributions have the same source) is still strongly rejected, with an average p-value of 1.55e-5. This is partly due to the sensitivity of the KS test to minute differences, but upon further inspection, this result can largely be attributed to the slight leptokurtosis exhibited (kurtosis~3.85) in the EEG data, which the Gaussian-based model (kurtosis~3) cannot replicate. This subtle shape mismatch, particularly in the tails, likely drives the low p-values despite strong similarity in other domains like autocorrelation and power spectral density. However, the qualitative observation of a close (but still imperfect) match is slightly reinforced by the p-value of the EEG estimate's KS test being significantly higher than that of the EMG estimate. In addition, the first, second, and third order statistics of the EMG estimate (mean = 3.3e-8, standard deviation = 4.44e-6, skew = -0.04) and real EMG data (mean = 3.9e-9, standard deviation = 4.36e-6, skew = -0.036) are very similar, indicating a close fit between the two distributions.

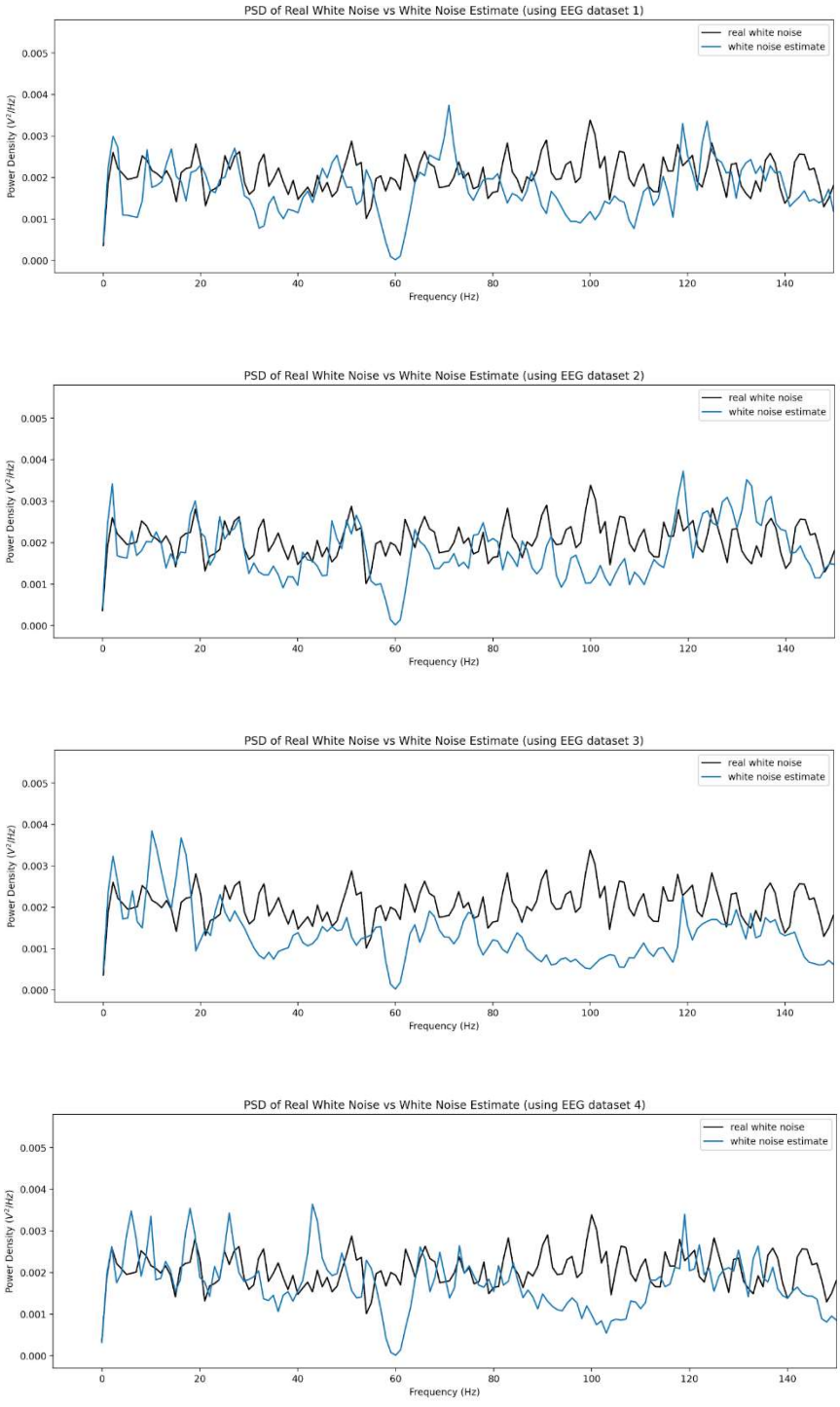


Figure 4-23. *Power spectral densities of white noise resulting from feeding each EEG dataset into the EEG model that was built using EEG dataset 5.*

At this point, it is also appropriate to evaluate the EEG model's generalizability across datasets. Following the same steps discussed at the end of section 4.2, the spectral flatness of the white noise that resulted from feeding each EEG dataset into the EEG model's inverse were as follows: 0.896 , 0.898 , 0.875 , and 0.886 (for EEG datasets 1, 2, 3, and 4, respectively). Compared to the original performance of 0.905, the model does not do much worse for EEG datasets 1 and 2, but a significant drop-off in performance can be observed in datasets 3 and 4 (Figure 4-23). Errors are particularly pronounced in the 80 to 120 Hz bandwidth, though another consistent source of error can also be observed in the 30 to 40 Hz bandwidth. However, a roughly uniform power distribution is maintained throughout most of the relevant bandwidths in each EEG dataset.

As with the EMG model, the generalizability of this EEG model is largely dependent on the level of variability in the EEG datasets. Because the frequency spectrum of brain activity is generally more complex and variable than that of EMG data - which is reflected in the variability in model performance for each EEG dataset - future research in this field may benefit from a focus on individualized models as opposed to generalized EEG models.

5. EEG-EMG Transfer Function

As mentioned in the introduction of this thesis, the goal of this research is to parameterize the relationship between EEG data and EMG data during prescribed movement. In the previous chapter, EEG and EMG data were parametrically characterized on their own, using individual, well-fitted autoregressive (AR) models. Through the methods discussed in this chapter, these characterizations can also be utilized to parameterize the EEG-EMG relationship, thereby providing a rough definition of the electrophysiological system that connects brain to muscle during squeezing movements.

5.1 EEG-EMG Parameterization Method

From Figures 32 and 41, we can see that AR signal models can be understood as all-pole filters that accept white noise with unit variance and return a signal with statistical characteristics similar to a desired signal. In addition, this all-pole filter representation of EEG/EMG signals also enabled performance evaluation by analyzing the spectral flatness of the "white noise" output of the inverse filter when driven by the desired signal. Given these two signal models and their inverse filter representations, we can connect EMG data to EEG data by driving the EMG model with the white noise that results from the inverse EEG filter, which in turn is driven by EEG data. As seen in Figure 5-1, this results in a system that accepts EEG data as input and returns an EMG estimate as output.

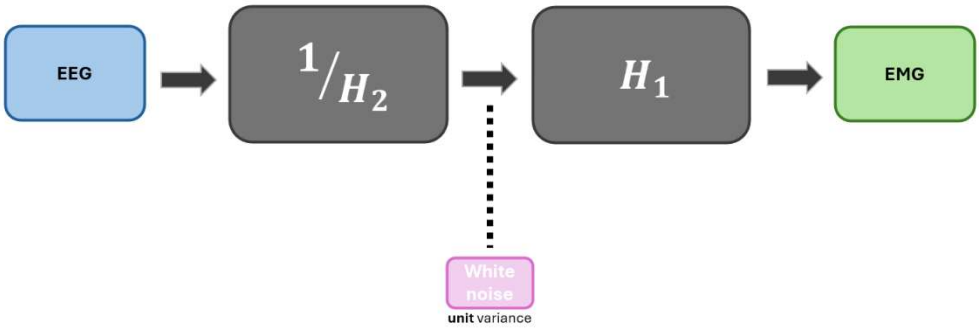


***Figure 5-1**. Parameterization of the EEG-EMG relationship through a synthesis of the individual EEG and EMG signal models.*

More concisely, the EEG-EMG relationship can be represented as a single system with a transfer function, as shown in Figure 5-2. This transfer function results from a straightforward combination of the inverse of the all-pole EEG filter (denoted $1/H_2$) and the all-pole EMG filter (denoted H_1), yielding a system with both poles and zeros. Notably, the b_0 of each individual model interact to create a scaling factor in the overall EEG-EMG system. This is an essential property in the EEG-EMG transfer function; as mentioned in Chapter 3, EEG signals generally exhibit amplitudes that are an order of magnitude lower than that of EMG signals, so any system that relates the two data sources should account for this discrepancy. However, as also discussed in Chapter 3, the power of squeeze-related EMG signals can vary across datasets, and this affects the b_0 scaling factor of the individual EMG model. Therefore, the exact value of this EEG-EMG scaling factor can vary depending on which EEG/EMG dataset the individual models are trained on.

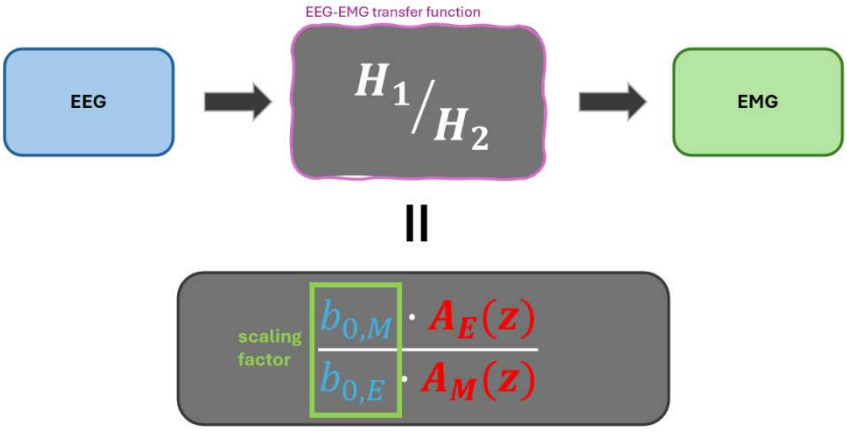


***Figure 5-2**. Succinct representation of the resulting EEG-EMG transfer function.*

Because this set-up assumes that the white noise coming out of the inverse EEG filter is just as spectrally flat as the white noise driving the EMG model in section 4.2, the whiteness of the white noise at the intermediate step in the process (as shown in Figure 5-1) is crucial to the EEG-EMG transfer function's performance. Additionally, the individual EEG/EMG models' assumption of stationarity is retained in this overall model; this transfer function relationship holds as long as both EEG and EMG data are wide-sense stationary. In the case of this thesis research, the transfer function can only be expected to perform as intended during squeezing motions. Although similar transfer functions can and should be developed using this method to define the EEG-EMG relationship during other movement-related tasks, this transfer function is not expected to accurately translate EEG data into EMG-like data in any other setting besides the experiment paradigm outlined in this thesis.

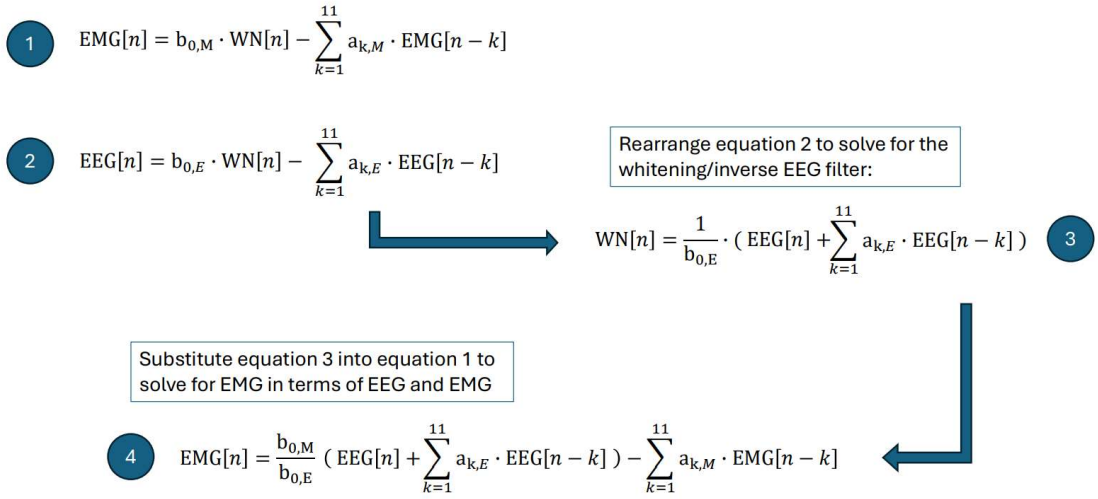


Figure 5-3. *Derivation of the EEG-to-EMG difference equation from individual autoregressive models of EEG and EMG data.*

Besides its treatment of EEG and EMG data as stochastic, non-deterministic data sources, the main benefits of this parameterization method can be found in its compressed definition of the EEG-EMG relationship. Through the use of autoregressive models, a set of filter coefficients can be used to define this complex, dynamic relationship. In the following subsections, it is shown that a set of less than 30 filter coefficients is often sufficient in defining this relationship. This efficient performance ability demonstrates a robust potential for lightweight, practical use in brain-computer-interface applications, as well as providing a condensed understanding of brain-muscle dynamics in electrophysiological pathways.

5.2 EEG-EMG Transfer Function: Model Selection

Because two separate AR models are involved in the construction of the overall EEG-EMG transfer function, order selection becomes slightly more nuanced than that of the individual

EEG/EMG models. Two primary approaches were considered.

The first approach uses the optimal EEG model and the optimal EMG model - which were both found to be AR(11) models in Chapter 4 - to build the overall EEG-EMG transfer function. This straightforward method has the benefit of adhering closely to the two-stage structure of the EEG-EMG transfer function: inverting EEG data into white noise, and then shaping pure white noise into EMG data. Using the individually optimal models in the overall transfer functions preserves the integrity and interpretability of the proposed EEG-EMG framework by providing emphasis on making sure that each stage performs its transformation as effectively as possible.

The second approach calculates an optimal transfer function from scratch, instead of using the individually optimal models selected in Chapter 4. This method involves generating a two-dimensional grid in which the normalized spectral error resulting from the EEG-EMG transfer function is displayed with respect to the order size of the EEG AR model and EMG AR model used. This method mirrors the technique used to find the optimal EEG/EMG models, in which performance was evaluated in comparison to the model size. In contrast, because we are no longer aiming to model a signal driven by a simple white noise input, we cannot use the spectral flatness of the inverse filter output to evaluate each model, but normalized spectral error provides a close, related metric. Ideally, the optimal transfer function is found to be the model with the lowest overall order (i.e, lowest sum of EEG and EMG model orders) within a plateau in performance. While less strict in adhering to the two-stage modeling structure, this approach allows for compensatory interactions between the EEG and EMG models – spectral deficiencies in one stage may be offset by incidental corrections in the other.

5.3 EEG-EMG Transfer Function: Properties and Evaluation

In the first, more basic method of EEG-EMG parameterization discussed in section 5.2, an EEG-EMG transfer function is formed from a simple combination of the individually optimal EEG and EMG models which are presented in Chapter 4.

As the EEG-EMG transfer function contains both poles and zeros, it is important to check that this transfer function is stable with minimum-phase. The stability of the EEG-EMG transfer function is ensured by the stability of the EMG model used, as the all-pole EMG model is the source of all of the poles that are present in this transfer function. Additionally, because the EEG AR model evaluated in section 4.3 is stable (all poles inside the unit circle), it can also be assumed that its inverse is minimum-phase (all zeros inside the unit circle). In other words, provided that the EEG model used in the EEG-EMG transfer function is stable, the EEG-EMG transfer function will act as a minimum-phase filter. As seen in Figure 5-4, all of the poles and zeros of the EEG-EMG transfer function are inside of the unit circle, confirming that the transfer function is both minimum-phase and stable. This is further demonstrated by the convergence of the impulse and step responses of the filter in Figure 5-5.

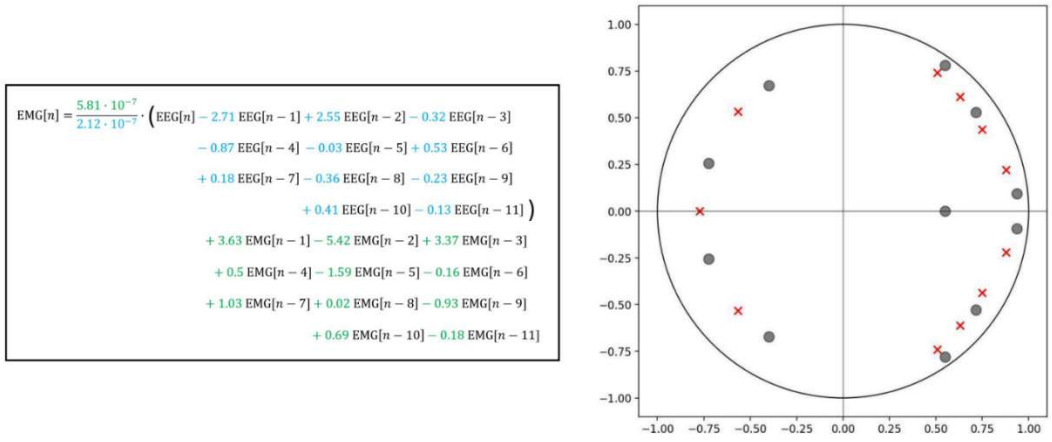


***Figure 5-4**. Difference equation (left) and pole-zero plot (right) of the 22-parameter EEG-EMG transfer function derived from the EEG and EMG models calculated in Chapter 4.2 and Chapter 4.3.*

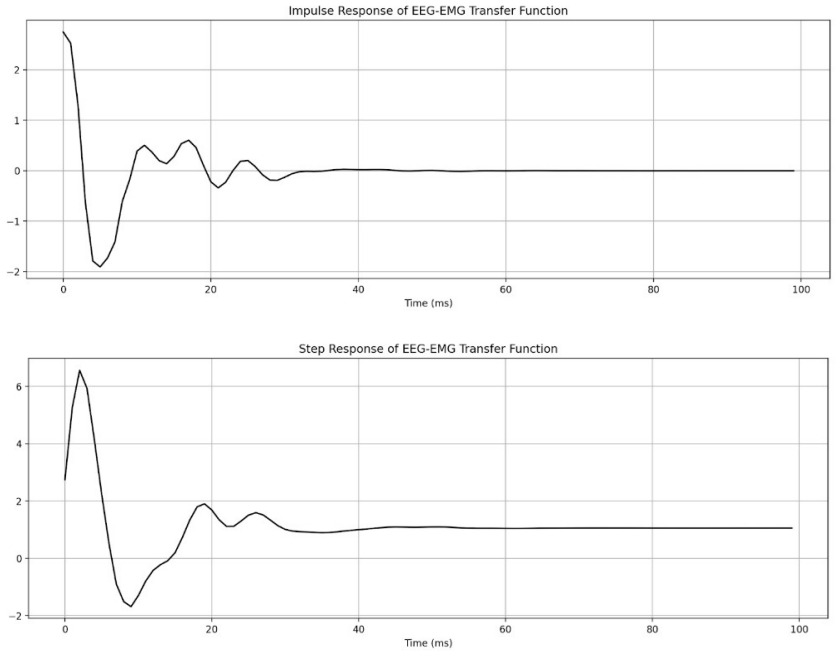


***Figure 5-5**. Impulse response (top) and step response (bottom) of the 22-parameter transfer function.*

The scaling factor of the EEG-EMG difference equation (presented in Figure 5-4) comes from the ratio of the scaling factors from the individual EEG and EMG models. As discussed in Chapter 4, these individual scaling factors are quite small due to the need to transform the unit power white noise input into noise with amplitudes similar to that of EEG and EMG waveforms. However, the ratio of these two scaling factors results in an overall scaling factor of approximately 2.74, which indicates that the transfer function scales the power of the input up. This matches our understanding that EEG waves generally have a smaller amplitude than that of EMG data - any function that relates EEG and EMG must account for this power mismatch.

To evaluate the performance of this transfer function (fitted using the squeeze-period data from EEG/EMG dataset 5), EEG data was fed into the EEG-EMG transfer function, and the

resulting EMG estimate was compared to the corresponding EMG data in a multitude of ways mirroring those presented in Chapter 4.

From the comparison of the EMG estimate and the EMG data's autocorrelation functions in Figure 5-6, it can be seen that this EEG-EMG transfer function scales the input EEG signal's power into EMG power levels quite well. Additionally, the autocorrelation of the EMG estimate matches that of the real EMG data closely up until approximately 10-11 ms of lag, where the subsequent slight dip in the EMG data's autocorrelation is overshot by that of the EMG estimate. This is in contrast to the autocorrelation of the EMG estimate presented in Chapter 4, which followed the autocorrelation of the EMG data for a few more milliseconds of lag. Additionally, oscillatory components in the later lag regions are more prominent in the EMG estimate's autocorrelation than in the autocorrelation of the EMG data, which smoothly converges to zero after 15 ms of lag.

Though initially concerning, this issue can be traced back to the absence of power in the EMG estimate's 57-63 Hz bandwidth, which in turn stems from the built-in EEG notch filter existing in the same region. Because the input to this EEG-EMG system (EEG data) has no power in this bandwidth, the transfer function is not able to transform the input in a way that produces any power in the same bandwidth - this results in the glaring mismatch in power spectral density in the 57-63 Hz region. This unavoidable property of EEG data also contributes to the unexpected oscillatory components in the autocorrelation of the EMG estimate, which oscillate at around 60 Hz. This can be confirmed by a visual inspection of the period of each oscillation, which ranges from roughly 15 to 18 ms, corresponding to frequencies between 55 and 65 Hz.

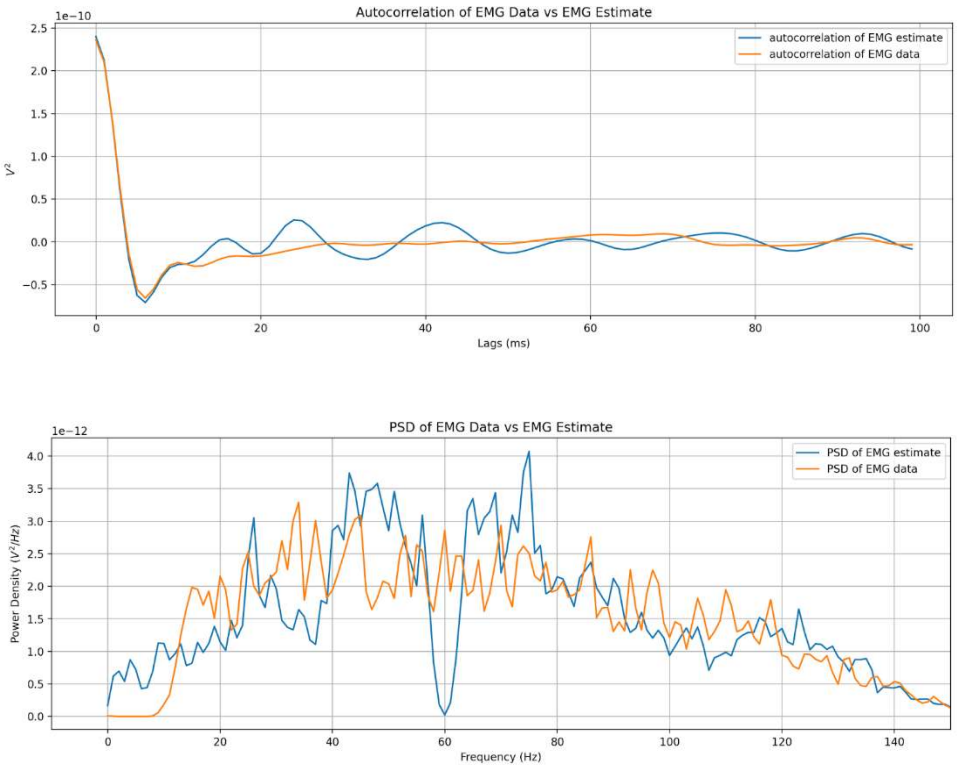


***Figure 5-6**. Comparison of the autocorrelation function (top) and power spectral density (bottom) of the EMG estimate and the real EMG data.*

Above 80 Hz, the power spectral density of the EMG estimate closely aligns with that of the actual EMG signal. A slight dip around 100 Hz is noticeable, which corresponds to a similar dip in the PSD of the white noise from the inverse EEG model (Figure 4-21). Similarly, power undershoots in the 30-40 Hz region and overshoots in the 45-50 Hz region are also reflected in the PSD of the white noise from the inverse EEG model, indicating that the inverse EEG model contributes more to the error of this EEG-EMG transfer function than the EMG model. As seen in Figure 4-11, the EMG model fits EMG data quite well - aside from the 10-20 Hz region - so the overall performance of the EEG-EMG transfer function largely depends on the accuracy of the inverse EEG model.

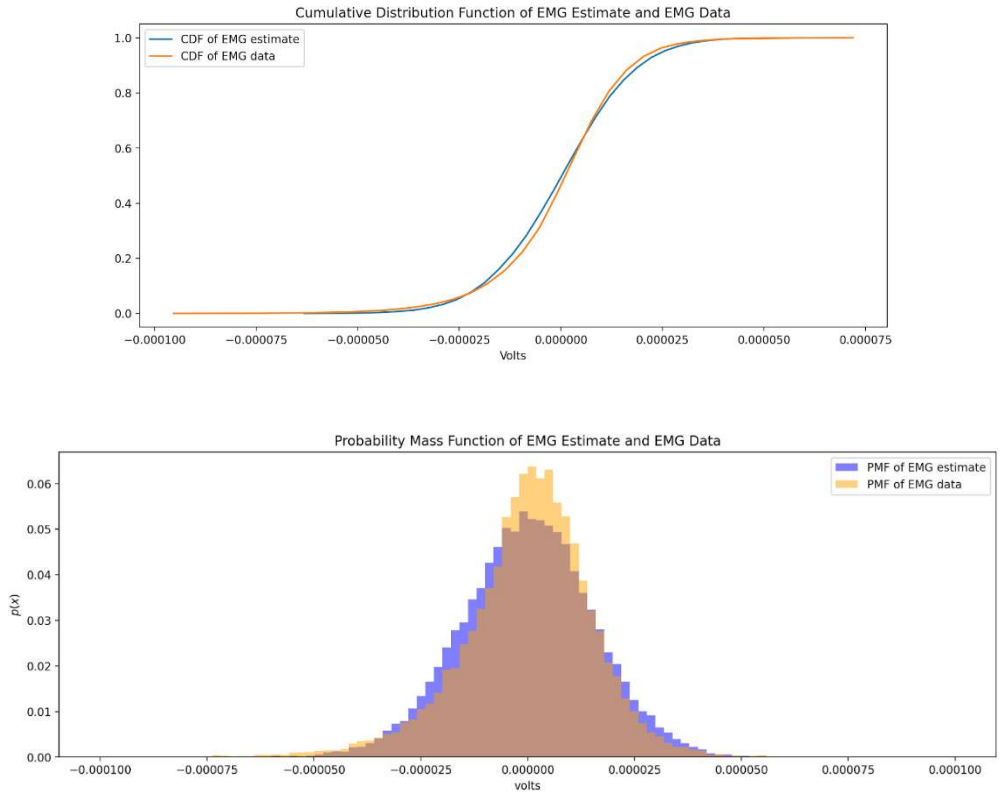


Figure 5-7*. Distribution of the EMG estimate compared to the distribution of real EMG data (EMG dataset 5).*

Barring a few mild aberrations, the PSD of the EEG-EMG transfer function's output has a similar overall shape to the PSD of the EMG data - power plateaus between 40 and 70 Hz, and then decreases smoothly as frequency increases. This shows that, given EEG data, the EMG estimate produced by the EEG-EMG transfer function closely approximates the shape of EMG data in the frequency domain.

The distribution of the EMG estimate resulting from the EEG-EMG transfer function is similar to the distribution of the EMG estimate from the individual EMG model discussed in section 4.2. This demonstrates that the skew, kurtosis, mean, and variance of the EMG estimate's distribution are not significantly affected as a result of using the white noise generated by the

inverse EEG model. However, a comparison of the distributions of the EMG estimate and EMG data shows that the same issues remain - the leptokurtic nature of this study's EMG data is difficult to replicate using EEG data as input, because EEG data has much less kurtosis in its distribution. The difference in kurtosis can be seen in the time-domain comparison (Figure 5-8), where EMG data shows a stronger central concentration and more frequent, pronounced spikes than the EMG estimate.

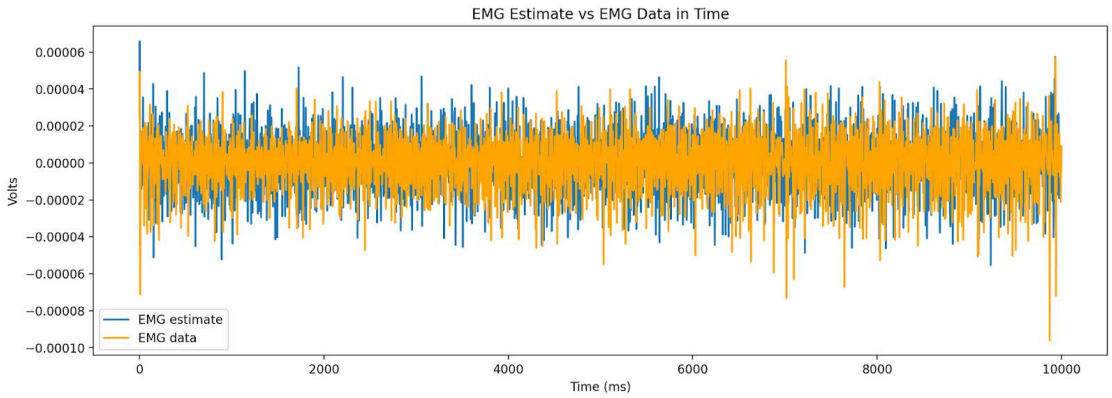


***Figure 5-8**. Time-domain comparison of the EMG estimate and EMG data during a squeeze/movement.*

Because the task of parameterizing the EEG-EMG relationship is best framed as a spectral matching task, a frequency domain-based analysis is appropriate in quantifying the performance of the EEG-EMG transfer function. However, the spectral flatness measure used in the evaluation of EEG and EMG signal models cannot be used to evaluate the performance of this EEG-EMG transfer function. Because the EEG and EMG signal models are driven by white noise, the spectral flatness metric is able to inverse the signal model/filter and evaluate it by measuring the whiteness of the inverse filter's output using a simple data-flatness measure. In contrast, the EEG-EMG transfer function is driven by EEG data, so inverting the transfer

function representation does not result in white noise and the straightforward data flatness measure cannot be used. As an alternative, the difference between the power spectral densities of the EMG estimate and the actual EMG data can be quantified using the normalized spectral error metric (NSE).

NSE (equation in Figure 5-9) quantifies the frequency-domain difference between the PSD of the EMG estimate and that of the actual EMG data by computing the average of the normalized absolute differences across all frequency bins. For example, if the EMG estimate has a power density of 2 V^2/Hz at 80 Hz and the actual EMG data has a power density of 2.5 V^2/Hz, then the normalized error in that frequency bin would be |2-2.5|/2.5 = 0.2). This normalized metric was chosen to avoid an unwarranted focus on high-power bandwidths, as normalization ensures that all frequencies contribute equally to error. Additionally, the L1 distance (absolute value) was chosen to avoid the skewing of performance evaluation from high amplitude outliers (as is common in L2-based metrics). It is not necessary for the PSD of the EMG estimate to be tightly fitted to that of the EMG data. As can be seen in the spectrograms of EMG data, exact peaks and dips in the EMG power spectral density fluctuate randomly; the more important pattern to model is the general shape of the EMG PSD.

Figure 5-9 shows the absolute error of the EMG estimate's PSD (resulting from the previously discussed 22-parameter EEG-EMG transfer function), plotted in red. This plot illustrates the EMG estimate's performance with frequency and quantifies our observation that the EMG estimate's performance in the frequency domain is better in frequencies above 80 Hz when the two previously discussed EEG and EMG models are used. One limitation of the NSE metric is that it averages the normalized absolute error across all frequency bins (from 10 Hz to 150 Hz), making it difficult to tell which frequencies contribute more to the overall error.

However, since the goal is to select the transfer function with the best overall spectral performance, the NSE metric remains appropriate for this task.

As discussed in section 5.2, the NSE metric can be used to determine the combination of EEG and EMG models that produces the most accurate transfer function - the jointly optimal transfer function. The results of this second approach (later referred to as the "grid method") are presented in Figure 5-10. EEG and EMG model orders were varied from 5 to 17, and the NSE was computed for each resulting transfer function. This procedure was repeated for each of the five EEG/EMG datasets, resulting in five 13x13 grids.

$$NSE = \frac{1}{f_2 - f_1} \sum_{f=f_1}^{f_2} \left| 1 - \frac{\hat{S}_{xx}(f)}{S_{xx}(f)} \right|$$

normalized error in the EMG estimate's power spectral density

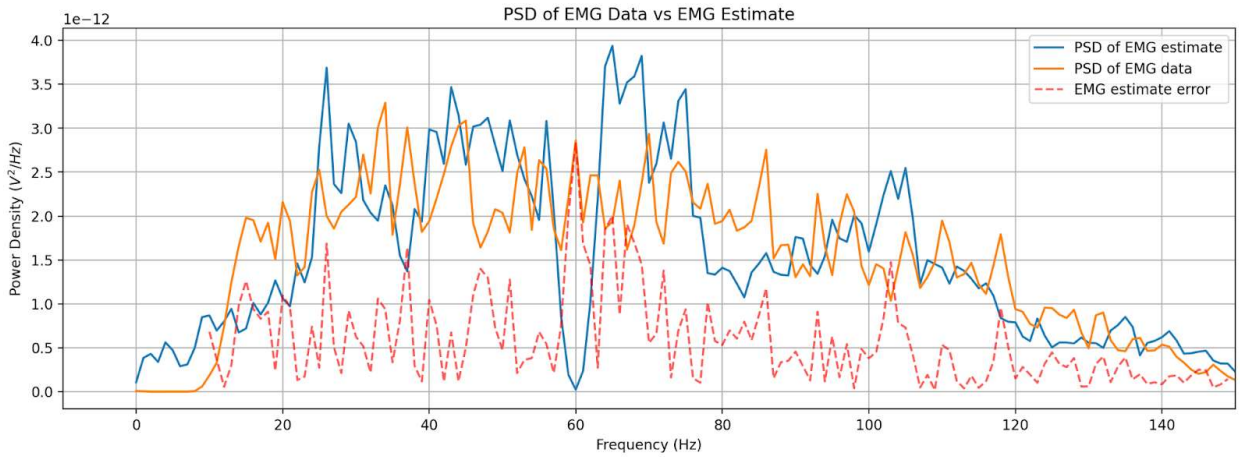
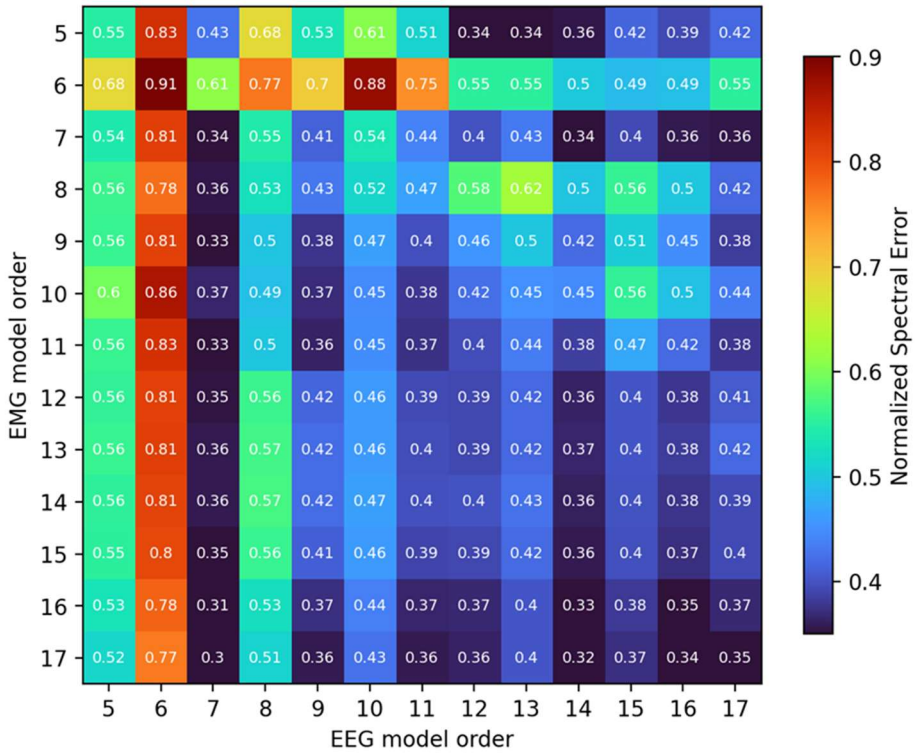


***Figure 5-9**. Top) Formula for normalized spectral error. Bottom) Comparison of EMG data and EMG estimate in power spectral density, with PSD error shown.*

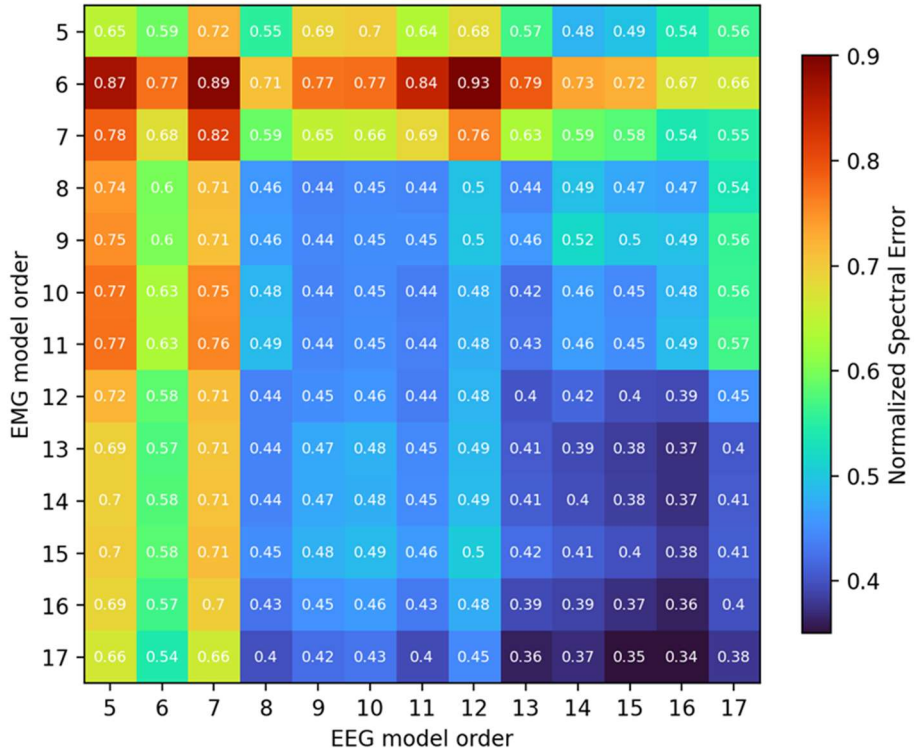
As seen in the NSE grid for EEG/EMG dataset 5, the normalized spectral error for the 22-parameter transfer function calculated using the individually optimal EEG and EMG models (11 poles for each model), is 0.38. This unexpectedly high magnitude of error can partly be attributed to the random fluctuations in the power spectral densities of the EMG data and the EMG estimate; even an ideally fitted estimate would have a considerable amount of incidental error. Therefore, more benefit can be derived from comparing the NSE of a certain transfer function to that of other transfer functions.

This choice of model order seems to be locally optimal, as adjacent transfer functions [*EMG*(11), *EEG*(12)] and [*EMG*(12), *EEG*(11)] have slightly higher normalized spectral error than the [*EMG*(11), *EEG*(11)] model, despite having more parameters. Although the [*EMG*(11), *EEG*(11)] transfer function performs well, we can see from this NSE grid that the transfer function is not the most optimal EEG-EMG transfer function for this dataset - it does not yield the best accuracy for the fewest number of parameters. Transfer functions [*EMG*(7), *EEG*(9)] and [*EMG*(11), *EEG*(8)] are a couple of models that yield slightly less NSE using fewer parameters. This is an example of one benefit of the grid method - it can help find transfer functions that achieve equivalent or better performance with fewer parameters (above the diagonal dotted line in Figure 5-11).

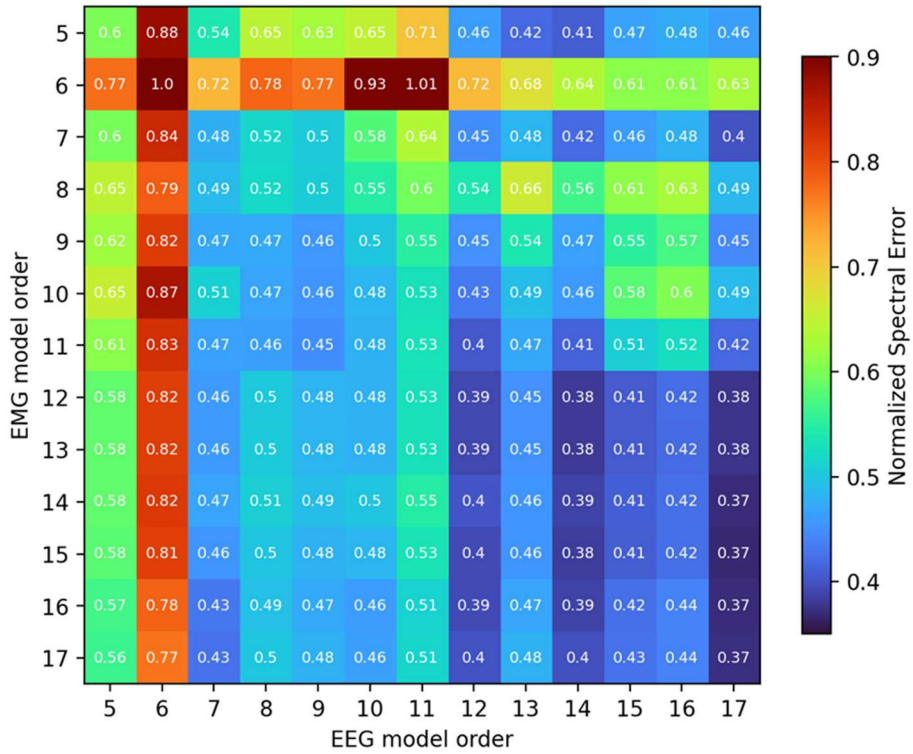
Normalized Spectral Error in EEG-EMG Transfer Function, Dataset 1



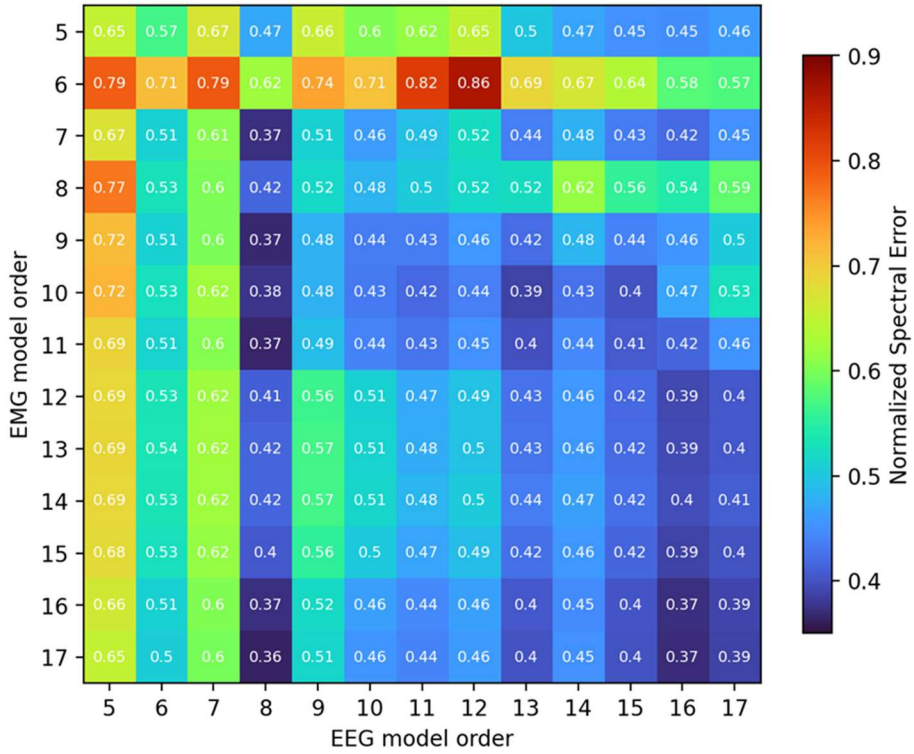
Normalized Spectral Error in EEG-EMG Transfer Function, Dataset 2



Normalized Spectral Error in EEG-EMG Transfer Function, Dataset 3



Normalized Spectral Error in EEG-EMG Transfer Function, Dataset 4



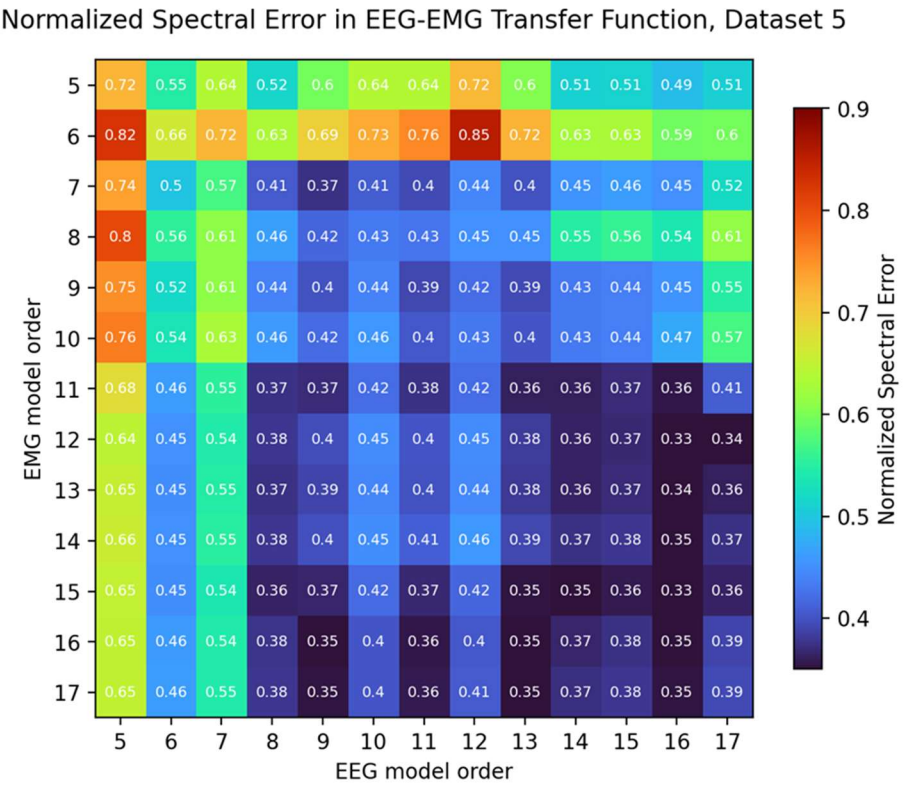


***Figure 5-10**. Results of evaluating transfer function performance as individual EEG and EMG model orders are varied, for each EEG/EMG dataset.*

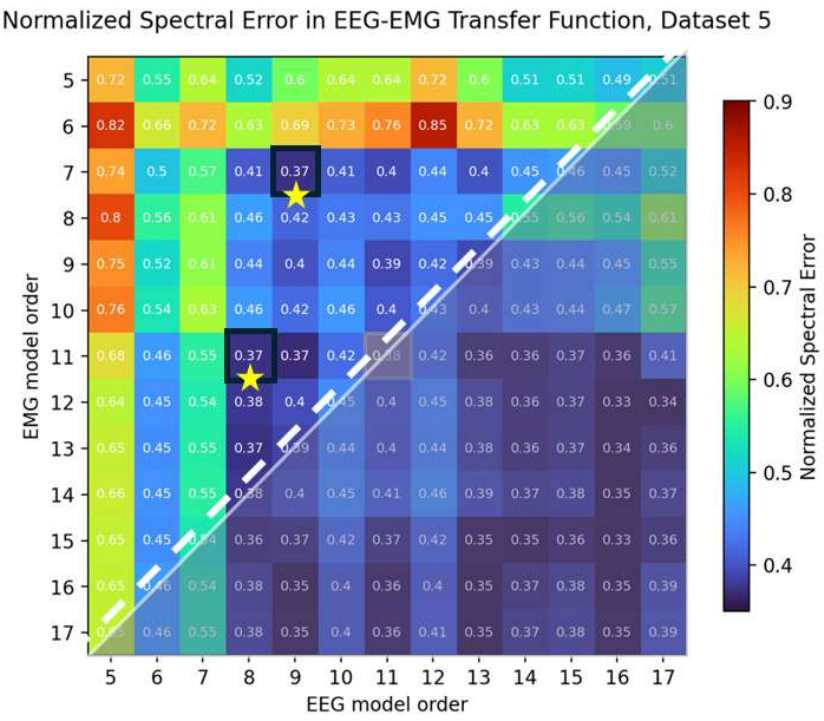


Figure 5-11*. Example of using the grid method to find more optimal transfer functions, which can be found in the upper triangular matrix of this grid.*

The power spectral densities of the EMG estimates that result from the [*EMG*(7), *EEG*(9)] and [*EMG*(11), *EEG*(8)] transfer functions (shown in Figures 61 and 62) demonstrate the compensatory interactions between EEG and EMG models that the grid method allows for.

The EEG(8) model used in the [*EMG*(11), *EEG*(8)] transfer function clearly overshoots the power of the EEG data in frequencies below 10 Hz, while undershooting the power in frequencies between 10 and 20 Hz. This results in a whitening filter that does the opposite - it provides less power below 10 Hz and more power between 10 and 20 Hz. When combined with the EMG model, which overshoots power in the 0-10 Hz range and undershoots power in the 10-20 Hz range, this results in a correction that provides much improved accuracy in the 10-20 Hz range of the EEG-EMG transfer function's EMG estimate, as seen at the bottom of Figure 5-12.

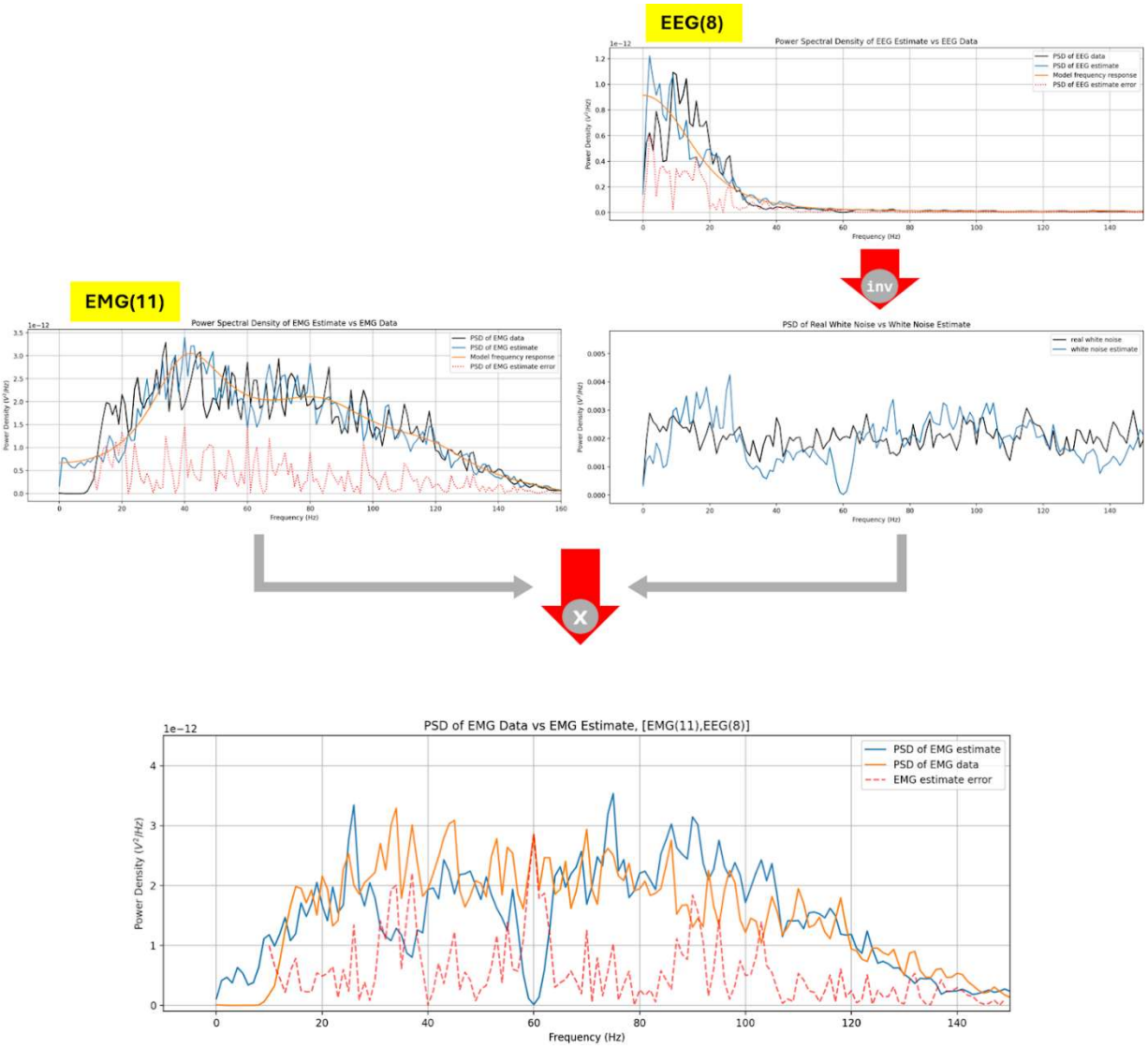


***Figure 5-12**. Frequency-domain performance of alternative transfer function [EMG(11), EEG(8)], for EEG/EMG dataset 5.*

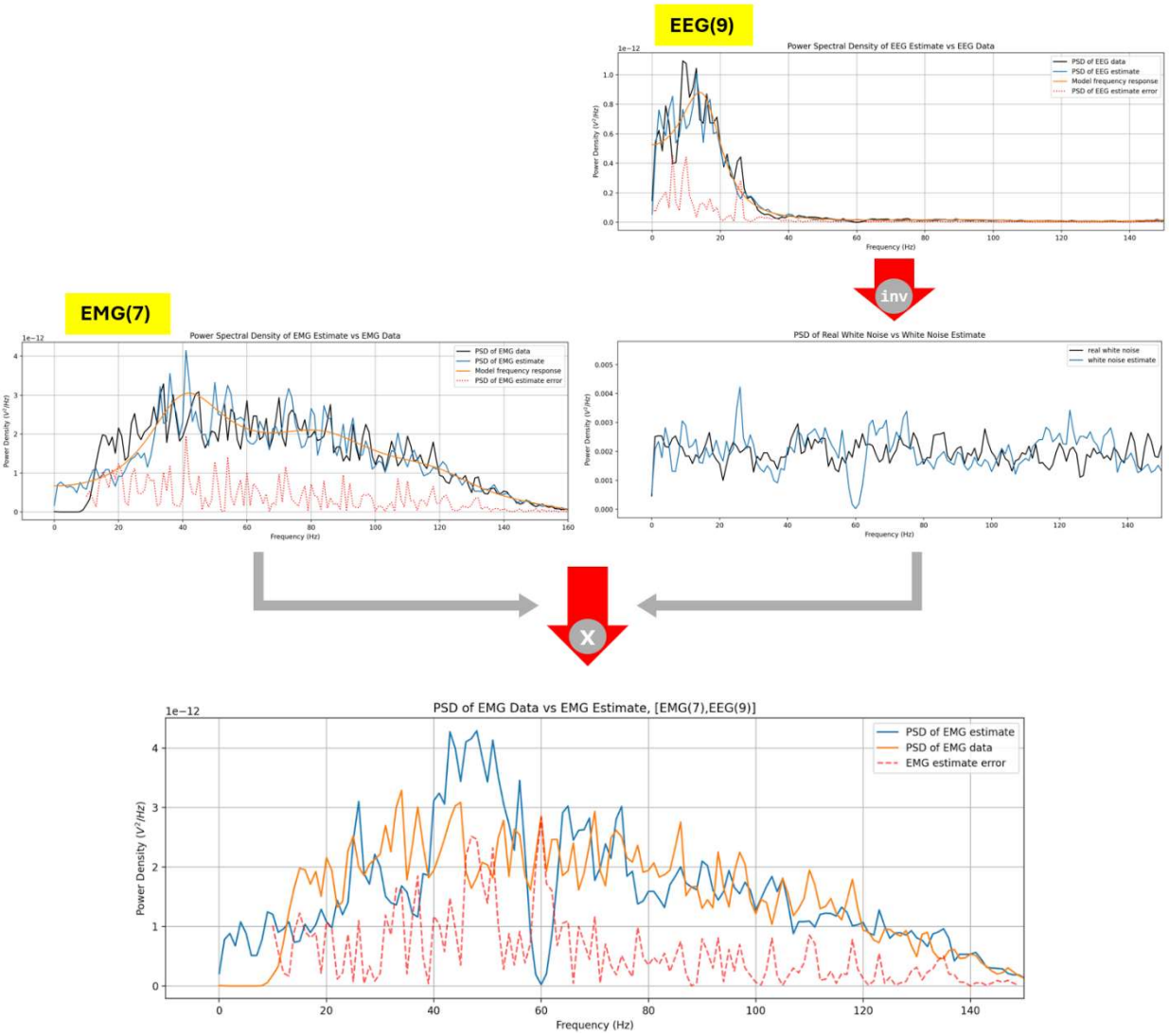


Figure 5-13. *Frequency-domain performance of alternative transfer function [EMG(7), EEG(9)], for dataset 5.*

The [*EMG*(7), *EEG*(9)] transfer function, on the other hand, uses an EEG model whose inverse more accurately filters EEG into white noise. While this has the benefit of making the 30-40 Hz dip in the transfer function's EMG estimate less pronounced than in the previous transfer function estimates, it also exacerbates a power overshoot between 40 and 55 Hz, caused by the EMG(7) model's frequency response over-amplifying frequencies between 35 and 60 Hz.

The tradeoffs between balancing performance in different frequency regions present a challenge in the EEG-EMG transfer function model selection. The straightforward combination of the individually optimal EEG and EMG models does better at ensuring a more balanced performance, whereas choosing a transfer function based on the grid method may result in a transfer function that has uneven performance in frequency.

The grid method also tells us more about what model order is required to sufficiently translate between EEG and EMG data. Concerning the EEG/EMG datasets used in this study, this threshold is often more dependent on the individual requirements of the EEG/EMG models than on the overall number of parameters in the transfer function. From the NSE grid for dataset 5, we can also observe a sharp jump in performance when moving to an EEG model order of 8, which indicates that having more than 7 poles in the EEG model is needed to adequately whiten the EEG signal in this dataset. Similarly, a sharp increase in performance can be seen as the EMG model goes from a model order of 6 to 7. Increases beyond this model order usually result in slow, sometimes inconsistent increases in performance.

Though the NSE grids presented in Figure 5-11 provide useful information on the dynamics of the interaction between the EEG and EMG models in the overall EEG-EMG transfer function, they also demonstrate the uniqueness of each dataset; the performance plateau doesn't happen at the same model orders each time, and even the plateau itself has varying characteristics. For example, some plateaus are defined by consistent, marginal increases in performance with model order, where others have certain model orders that have markedly superior performance than adjacent model orders. Some datasets may not have a clearly distinguishable plateau at all.

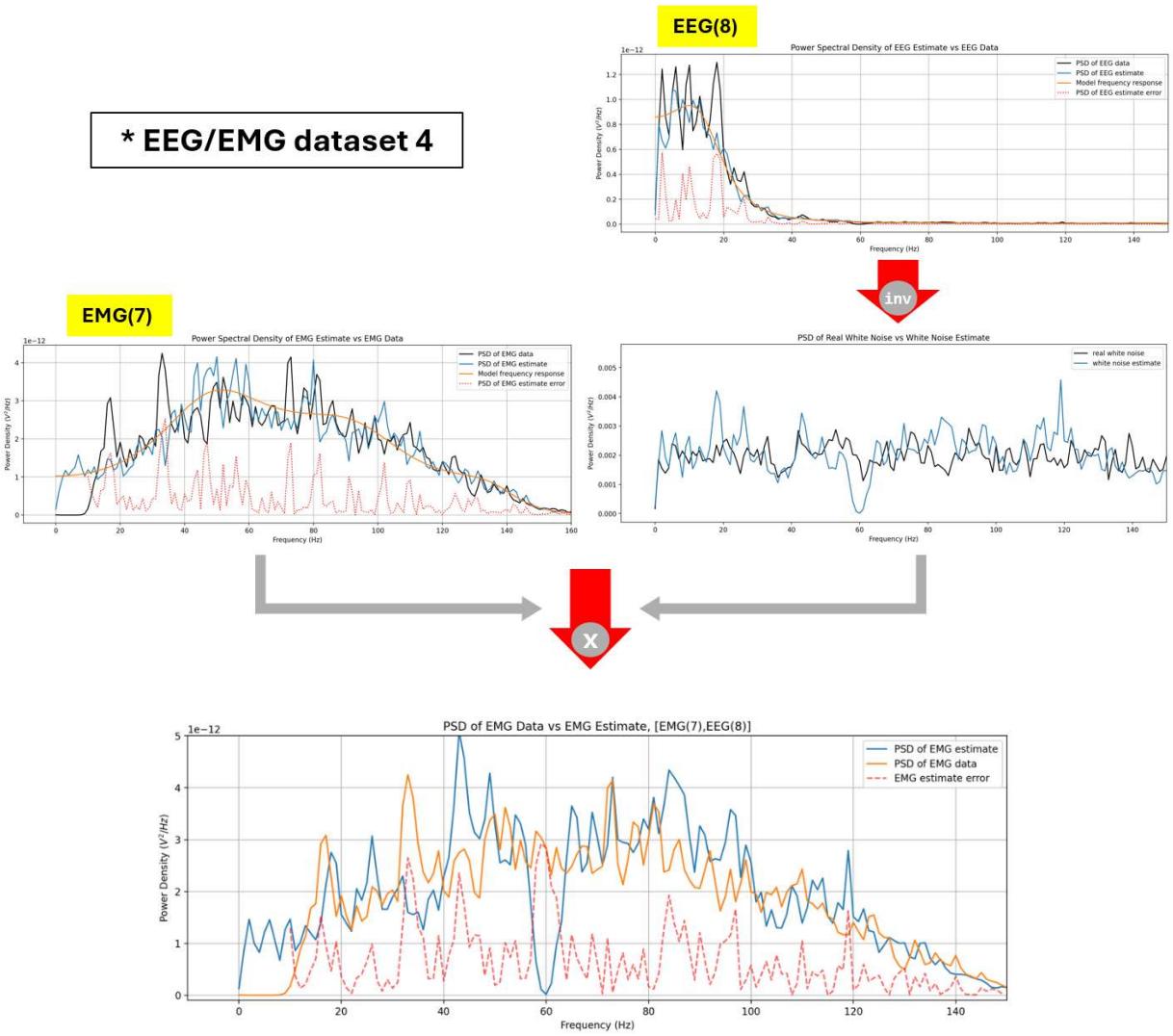


Figure 5-14. *Frequency-domain performance of the optimal transfer function [EMG(7), EEG(8)] for dataset 4.*

A prime example of this uniqueness is the NSE grid for EEG/EMG dataset 4. Immediately noticeable is a column of markedly excellent performance corresponding to an EEG model order of 8 (especially great for EMG model orders 7, 9, 11). This is an extremely potent finding, as knowing that the [*EMG*(7), *EEG*(8)] model performs just as well as transfer functions with over 30 parameters indicates a strong opportunity for a more efficient transfer function. Figure 5-14 demonstrates that this excellent performance comes partly from the shape of EEG

dataset 4 being easier to model, especially in the 2-15 Hz range. Additionally, the EMG data has a stronger peak at 50 Hz than dataset 5, which makes it a bit easier for the EMG data to be modeled - AR models are best suited to modeling signals that have peaks in frequency, because AR models are all-pole filters, and poles correspond to frequency peaks. This characteristic of the NSE grid for EEG/EMG dataset 4 is not perfectly replicated in any other datasets; EEG/EMG dataset 1 has a similar column of stand-out performance, but it is related to an EEG model order of 7, not 8. Other characteristics, like the normalized spectral error that the EEG-EMG transfer functions converge to (in the performance plateaus), also differ from NSE grid to NSE grid.

Although the performance of the EEG-EMG transfer functions across different EEG/EMG model orders is not identical across the five EEG/EMG datasets, the NSE grids in Figure 5-11 indicate that EEG-EMG transfer functions should generally have EEG models with more than 6 poles and EMG models with more than 5 poles (in most cases) to effectively translate from EEG data into EMG data. When this is the case, optimal EEG-EMG transfer functions typically result in normalized spectral error near 0.37, and qualitative analysis reveals the compensatory interactions between the individual EEG and EMG models contribute to slightly improved performance over the more basic method of combining individually optimal EEG and EMG models. However, the straightforward method may be used in cases where a more generalizable and interpretable EEG-EMG transfer function is desired.

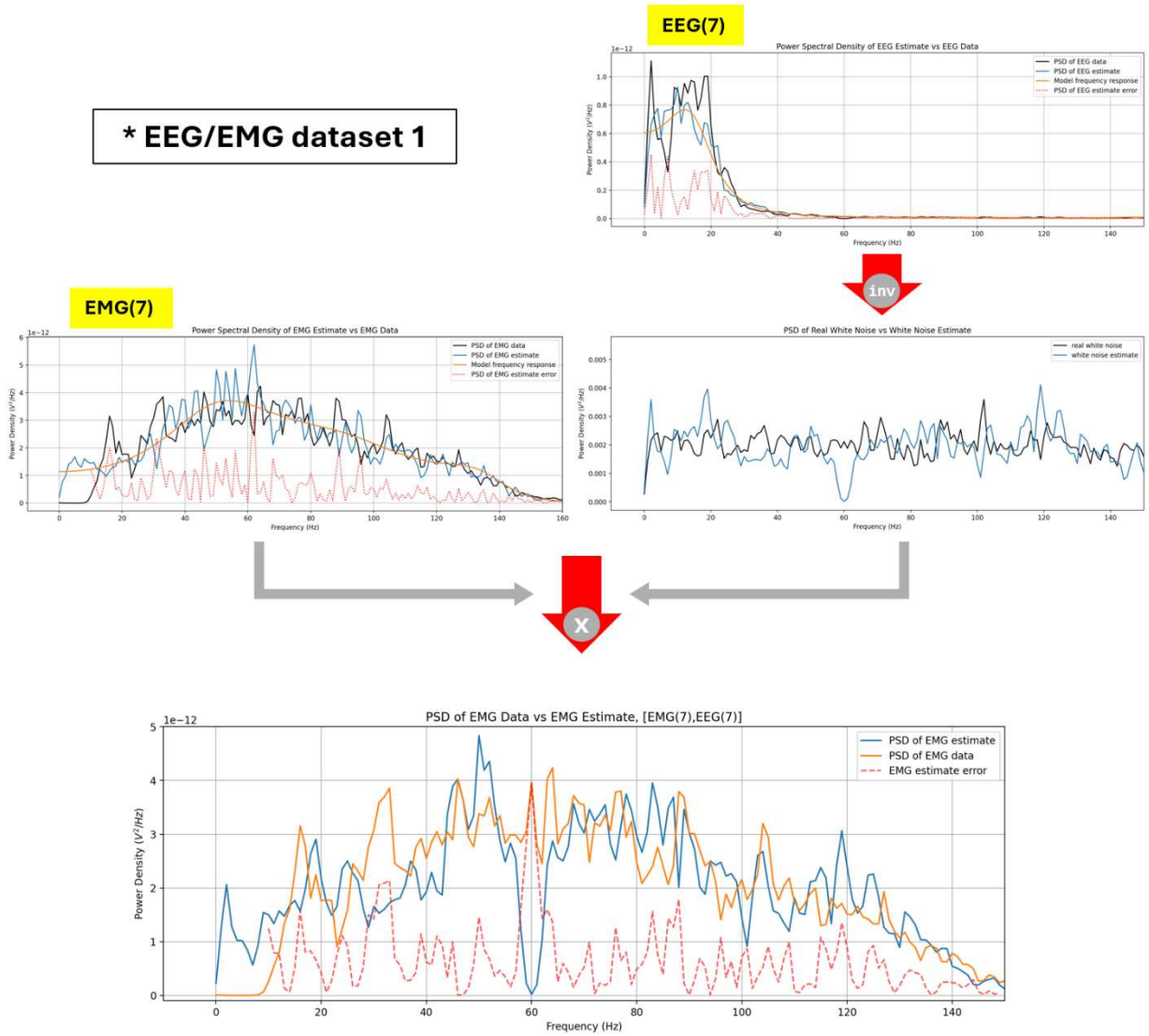


***Figure 5-15**. Frequency-domain performance of the optimal transfer function [EMG(7), EEG(7)] for dataset 1.*

6. Conclusions and Future Work

This thesis presented an analysis of EEG and EMG data during prescribed movement (a grip-squeezing task), modeled the EEG and EMG data using linear predictive autoregressive models, presented a method for modeling the relationship between EEG and EMG signals using a transfer function derived from the AR models, and evaluated these transfer functions using normalized spectral error. By approximating the systems that generate EEG and EMG data as linear processes and implementing the ratio of their respective models (H1/H2), a novel EEG-EMG transformation was constructed. This approach enables a parameterizable system-level view of the brain-muscle connection via a data-driven framework that prioritizes interpretability.

Given the respective EEG data, the resulting EMG estimates generated by the EEG-EMG transfer function demonstrated both qualitative and quantitative resemblance to EMG in terms of both first and second order statistics, including power spectral density and autocorrelation. Models were found to always be stable with minimum-phase, as ensured by the calculation method used to determine the AR models used in the transfer functions. Spectral flatness and normalized spectral error were used to demonstrate performance in both the individual EEG and EMG models as well as the overall EEG-EMG transfer function.

A few limitations of the EEG-EMG transfer function remain. The distributions of EMG estimates generated by the EEG data and the EEG-EMG transfer function exhibited kurtoses that were consistently lower than that of the real EMG data, highlighting the model's inability to transform third and fourth order elements of the data's distribution. To tackle this issue, normalization of each distribution could allow for an unbiased evaluation of the distribution transformation involved in the EEG-EMG transfer function.

Additionally, built-in filters in both EMG data (low-pass filter) and EEG data (notch filter) significantly interfered with the model's calculated frequency responses and negatively skewed the spectral flatness/normalized spectral error results. Future studies could benefit from the use of selective linear predictive modeling, in which only certain spectrums of the data in question are fitted (Makhoul, 1975).

Future use of EEG-EMG transfer functions as presented in this thesis range from brain-computer interfaces to a more clinical, diagnostic use. Just as the poles in EEG AR models have been used as features to diagnose Parkinson's disease, the poles and zeros of optimal EEG-EMG transfer functions can be used to characterize an individual's brain-muscle connection, perhaps during a particular movement. Studies may find value in investigating the differences in the optimal transfer functions that result from variations in applied force, neuromuscular health, or fatigue levels. Ultimately, EEG-EMG transfer functions, as presented in this thesis, provide an additional, meaningful basis on which to parameterize and define the connection between electrical activity in the brain's motor cortex and electrical activity in the forearm muscles during muscle contraction.

7. References

1. J. Makhoul, "Linear prediction: A tutorial review" in Proceedings of the IEEE, vol. 63, no. 4, pp. 561-580, April 1975, doi: 10.1109/PROC.1975.9792.

2. Nayak CS, Anilkumar AC. “EEG Normal Waveforms.” [Updated 2023 Jul 24]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK539805/

3. Schwartz AB. “Movement: How the Brain Communicates with the World.” Cell. 2016 Mar 10;164(6):1122-1135. doi: 10.1016/j.cell.2016.02.038. PMID: 26967280; PMCID: PMC4818644.

4. Brain Products GmbH. (2008). ActiCAP Active Electrode EEG System [Apparatus and Hardware]. Brain Products. https://www.brainproducts.com

5. Delsys Inc. (2017). Trigno™ Quattro Wireless EMG Sensor [Hardware]. Natick, MA: Delsys Inc. https://www.delsys.com/product/trigno-quattro/

6. Enders H, Nigg BM. “Measuring human locomotor control using EMG and EEG: Current knowledge, limitations and future considerations.” Eur J Sport Sci. 2016;16(4):416-26. doi: 10.1080/17461391.2015.1068869. Epub 2015 Aug 4. PMID: 26238032.

7. Liu J, Sheng Y, Liu H. “Corticomuscular Coherence and Its Applications: A Review.” Front Hum Neurosci. 2019 Mar 20;13:100. doi: 10.3389/fnhum.2019.00100. PMID: 30949041; PMCID: PMC6435838.

8. Keihani A, Mohammadi AM, Marzbani H, Nafissi S, Haidari MR, Jafari AH. “Sparse representation of brain signals offers effective computation of cortico-muscular coupling value to predict the task-related and non-task sEMG channels: A joint hdEEG-sEMG study.” PLoS One. 2022 Jul 1;17(7):e0270757. doi: 10.1371/journal.pone.0270757. PMID: 35776772; PMCID: PMC9249190.

9. N. N. Tun, F. Sanuki and K. Iramina, “EEG-EMG Correlation Analysis with Linear and Nonlinear Coupling Methods Across Four Motor Tasks.” 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, 2021, pp. 783-786, doi: 10.1109/EMBC46164.2021.9629969.

10. Xi, Xugang & Ma, Cunbin & Yuan, Changmin & Miran, Seyed M. & Hua, Xian & Zhao, Yun-Bo & Luo, Zhizeng. (2020). “Enhanced EEG–EMG coherence analysis based on hand movements.” Biomedical Signal Processing and Control. 56. 101727. 10.1016/j.bspc.2019.101727.

11. M. Fauvet, S. Cremoux, A. Chalard, J. Tisseyre, D. Gasq and D. Amarantini, “A novel method to generalize time-frequency coherence analysis between EEG or EMG signals during repetitive trials with high intra-subject variability in duration.” 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER), San Francisco, CA, USA, 2019, pp. 437-440, doi: 10.1109/NER.2019.8716973.

12. Gao Y, Ren L, Li R and Zhang Y. (2018). “Electroencephalogram– Electromyography Coupling Analysis in Stroke Based on Symbolic Transfer Entropy.” Front. Neurol. 8:716. doi: 10.3389/fneur.2017.00716

13. G. Cisotto, A. V. Guglielmi, L. Badia and A. Zanella, "Classification of grasping tasks based on EEG-EMG coherence." 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom), Ostrava, Czech Republic, 2018, pp. 1-6, doi: 10.1109/HealthCom.2018.8531140.

14. Kim, Byeongnam & Kim, Laehyun & Kim, Yun-Hee & Yoo, Sun. (2017). “Cross-Association Analysis of EEG and EMG Signals According to Movement Intention State.” Cognitive Systems Research. 44. 10.1016/j.cogsys.2017.02.001.

15. Gwin JT and Ferris DP. (2012). “Beta- and gamma-range human lower limb corticomuscular coherence.” Front. Hum. Neurosci. 6:258. doi: 10.3389/ fnhum.2012.00258

16. Yuan H, Perdoni C, He B. “Relationship between speed and EEG activity during imagined and executed hand movements.” J Neural Eng. 2010 Apr;7(2):26001. doi: 10.1088/1741-2560/7/2/026001. Epub 2010 Feb 18. PMID: 20168002; PMCID: PMC3036745.

17. T. Bohlin, "Comparison of Two Methods of Modeling Stationary EEG Signals," in IBM Journal of Research and Development, vol. 17, no. 3, pp. 194-205, May 1973, doi: 10.1147/rd.173.0194.

18. Nazarpour K, Al-Timemy AH, Bugmann G, Jackson A. “A note on the probability distribution function of the surface electromyogram signal.” Brain Res Bull. 2013 Jan;90:88-91. doi: 10.1016/j.brainresbull.2012.09.012. Epub 2012 Oct 6. PMID: 23047056; PMCID: PMC3878385.

19. Chen, G. (2014). “Are electroencephalogram (EEG) signals pseudo-random number generators?” Journal of Computational and Applied Mathematics, 268, 1–4. https://doi.org/10.1016/j.cam.2014.02.028

20. Therrien, C. W. (1992). Discrete Random Signals and Statistical Signal Processing. Prentice Hall.

21. De Luca, Carlo J. “Surface Electromyography: Detection and Recording.” References - Scientific Research Publishing, DelSys Incorporated, 2002, www.delsys.com/downloads/TUTORIAL/semg-detection-and-recording.pdf.

22. Agarwal, Rakhi, et al. “EMG and grip force correlation at varying wrist angles for dominant vs. non-dominant hand.” 2018 IEEE International Students’ Conference on Electrical, Electronics and Computer Science (SCEECS), Feb. 2018, pp. 1–4, https://doi.org/10.1109/sceecs.2018.8546850.

23. Pew Research Center, March 2022, “AI and Human Enhancement: Americans’ Openness Is Tempered by a Range of Concerns.”

Image References

Figure 3-2:

Example of EEG use: https://www.inverse.com/science/eeg-100-years-brain-activity-neuroscience

Example of EMG use: https://www.medicalexpo.com/prod/delsys/product-123919-989579.html

Figure 3-3:

Delsys Trigno Quattro: https://www.medicalexpo.com/prod/delsys/product-123919-989584.html

Figure 3-13:

actiCap SNAP EEG system: https://pressrelease.brainproducts.com/easycap-cap-overview-2025/

actiCHAMP PLUS amplifier: https://brainvision.com/products/actichamp-plus/

Figure 3-15:

Left: https://www.learningeeg.com/montages-and-technical-components

Right: https://www.researchgate.net/figure/Average-reference-derivation_fig2_48408346

Figure 3-22:

Low frequency EEG bands: https://link.springer.com/chapter/10.1007/978-3-031-23602-0_11