Spectral Power and Mutual Information between 2 EMG Channels

Rujul Kapadia, B.Tech

A Dissertation

Presented to the University of Dublin, Trinity College

in partial fulfilment of the requirements for the degree of

Master of Science in Computer Science (Data Science)

Supervisor: Prof. Bahman Honari

August 2019

Declaration

I, the undersigned, declare that this work has not previously been submitted as an exercise for a degree at this, or any other University, and that unless otherwise stated, is my own work.

Rujul Kapadia,

August 14, 2019

Permission to Lend and/or Copy

I, the undersigned, agree that Trinity College Library may lend or copy this thesis upon request.

Rujul Kapadia,

August 14, 2019

Acknowledgments

I would like to take this opportunity to thank my parents, who have always supported me and motivated me.

I would also like to thank Prof. Bahman Honari and Prof. Bahman Nasseroleslami who have guided me all throughout my dissertation and without whose support this dessertation would not have been possible.

Lastly, I would thank my friends Ravi and Radhika who have always been by my side, as well as my classmates from whom I have learnt a lot.

Rujul Kapadia,

University of Dublin, Trinity College
August 2019

Spectral Power and Mutual Information between 2 EMG Channels

Rujul Kapadia, , Master of Science in Computer Science

University of Dublin, Trinity College, 2019

Supervisor: Prof. Bahman Honari

Measuring associations in motor neurons through electrophysiological monitoring methods like electromyography is integral to neuroscience as it helps in understanding the underlying synchrony between the muscles and the brain. This can be useful in the diagnosis of several neuromuscular and genetic disorders like the ALS and the Duchene's muscular dystrophy.

Information theory provides tools to estimate these interdependencies, of which non-linear measures like the mutual information can provide the best results, given to that it is the independent of the structure of data along with it's average being meaningful. Calculation of mutual information using the newly established density-based approach is performed in this study to understand the interactivity between two hand muscles. The results of these are validated against the traditional bin-based approach which requires much larger quantities of data to arrive at precise estimations. Finally, a comparison with the linear methods - the Power Spectrum and the Coherence are made to determine the overall performance of the used method.

The results of this dissertation can be used to conduct detailed studies in a wide range of neuroscientific experiments as much lesser quantities of data is required for studying the neural interactions than that of conventionally used. Moreover, these can also be employed to design advanced prosthetics.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The brain is a complex network of densely interconnected neural circuits that control our behavior. Understanding how these neural structures integrate and perform computation is integral for comprehending brain functions[5]

Since neural communications are governed by electrical signals, it is essential to record this data for research purposes. Electrodiagnostic and neuroimaging techniques capture such neuroelectric signals, of which, electromyography examines the interactivity between motor neurons that control the muscles in our body. Muscles contractions cause the nerve cells to transmit electrical signals which are translated by the EMG electrodes into interpretable formats like sounds, graphs or numerical values[4]. Results from this electrodiagnostic technique are then interpreted by specialists to understand the activation levels and detect complications in neuromuscular signal transmissions and dysfunctions in nerves and muscles. However, such neuroscientific experiments deal with data that are multivariate in nature and have nonlinearity in the variable interactions[8].

Information theory plays an essential role in this case and is highly equipped to contend to such data. It encompasses tools for multivariate analysis which can be asserted to various kinds of data. Moreover, it does not rely upon presumptions on the underlying structure of the data and can capture non-linearity in the interactions.[5]

This study makes use of one such tool: mutual information to understand the interdependency between muscular regions. Findings from this can be used to finely examine the muscular synchrony which can aid in the diagnosis of neuromuscular dis-

order, for conducting research in kinesiology, and for designing advanced prosthesis devices[4][prosthetic paper].

1.1 Motivation

Several information theory quantities can be used to measure the relationship between two EMG channels. However, simply using linear measures like coherence does not provide an accurate estimation of the relationship of the intermuscular activities. This happens so because such measures rely on phase relationships and power and a variation of any of these entities directly impacts their values. Moreover, other metrics for statistical dependence like Spearmans rho, or the coefficient relations by Pearson need an assumption of the marginal distribution. Therefore, non-linear measures like mutual information are considered as part of this study to measure the relationship between 2 muscles. Also, mutual information is a measure that can be summed up and thus, taking a mean of this quantity is meaningful.

Many approaches have been used to estimate the mutual information between two spike channels, like converting the spike trains into bins and calculating the mutual information based on the summed data points with respect to the bins. However, an enormous amount of data is required to calculate MI using such conventional methods. Additionally, precise evaluations from such methods require quantities of data that is practically infeasible to record. Therefore, instead of using estimation methods that rely on the dimensions of the data, an approach is chosen such that no coordinates are required for the calculation of mutual information. This ensures that a comparatively minimal amount of data is required to obtain an accurate estimation.

Apart from this, electrophysiological data like the EMG typically either have values that lie in an integrable manifold or is discrete. Though such problems in the estimation of mutual information have already been addressed [5, 9, 10], there havent been methods that address this difficulty in line with the above problem requiring enormous quantities of data. Therefore, the KozachenkoLeonenko estimator was used in this dissertation.

Thus, this study can be used to validate the newly established methods against the traditional approaches which can be further used to efficiently study motor control as well as used for the diagnosis of disrupted brain functioning and neuromuscular

disorders.

1.2 Research Question

How can estimations of mutual information between EMG channels using the density-based approach be used to quantify intermuscular relationship?

1.3 Research Objective

To address the research question of the dissertation, the following objectives are considered:

1. Using the innovative density-based method to estimate the mutual information between 2 EMG channels
2. Calculating the mutual information using the bin-based approach and cross-comparing the KL approach (or density-based approach) using simulated and experimental data
3. Using simulated data to understand the effect of parameters in the two approaches by employing variations in the count of bins and the smoothing parameter, h
4. Choosing an appropriate value of the smoothing parameter for the experimental data
5. Comparing the results of auto-mutual information of two EMG channels to their respective Power Spectrums
6. Analyzing mutual information between two EMG channels against the intermuscular coherence.

1.4 Research Challenge

1. Experimenting with the different values of the smoothing parameter with each test taking an enormous amount of time.

2. Choosing an optimal count of bins for the bin-based approach
3. Selecting the optimal value of the smoothing parameter, h
4. Difficulty in visually comparing a large number of MI estimates from the bin-based and density-based approach
5. Analyzing estimations of mutual information using different values of the smoothing parameter with some results coinciding each other.
6. Since the correlation between spike trains vary at a particular point in time varies, providing an overall estimate of mutual information between the 2 channels is difficult.

1.5 Thesis Overview

The density-based approach is a newly established procedure for the calculation of mutual information between electromyograph channels. This approach gives a precise estimation of the behavioral relationship of the muscular regions and accounts for the large quantities of data used in the earlier methods

The data for the study is granted by the Trinity Biomedical Sciences Institute. EMG data was taken for 8 subjects using a noninvasive method for the two-hand muscles - APB and FDI. With a sampling rate of 2048Hz, the data for each individual for 4 seconds and several such trials was present for each patient to account for any disruptions in recordings. Apart from this, simulated data is also used as part of this study which is created with the help of MATLAB. Furthermore, downsampling by a factor of 8 is performed on both the numerical and experimental data.

Hereafter, a comparison of the new approach is made to the conventional bin-based approaches that require the spike train to take values on metric space. Simulated Data is used to showcase how the performance of the chosen approaches gets effected with the variation of their respective parameters.

Finally, the two methods are compared to other measures: Coherence and Power Spectral Density using the experimental data. The decimation of the simulated and experimental data, as well as the implementation, is performed with the help of MATLAB.

1.6 Thesis Structure

Section 1 provides a brief introduction to the research and highlights the research question. Moreover, it includes a small discussion on the research objective, motivation, and the challenges faced during the project.

Followed by a literature review on several measures that quantify the relation between Neuroelectric signals, section 2 aims to provide background on EMG and their applications

The focus of Section 3 is on the two methods employed in this study. This section also includes a brief discussion of other information theory approaches.

Section 4 describes the data used in this study: Simulated and Experimental and provides an outline of how downsampling was performed.

Section 5 shows how the two methods were implemented and showcase the results of the various experiments that were carried out in the dissertation.

The conclusion of the dissertation is discussed in Section 6. Further, the results of the study are reviewed, limitations are discussed and a brief remark on the future work of this analysis is made.

Chapter 2

Related Work and Background Study

2.1 Estimations of interdependency between Spike trains using non-Frequency based paradigms

Several studies have been conducted that examine the spectral pattern in the spike trains of neurons. Many such approaches making use of the continuous electrophysiological data can be broadly classified into frequency-domain and time-domain methods. This section briefly discusses the temporal methods along with several others like the PSTH, JPSTH, Classification models and likelihood models.

Unnormalized shuffle-correlated cross-correlogram is a time-domain approach to detect the synchronization of the timing of the neuronal spiking [11]. Here, spike trains are first binned for a specified bin length. Thereafter, correlogram for several sequences of lags is calculated for the pair of spike data by using covariance between the binned pairs. An illustration of the covariance is given by the timing of the spiking activity, excitatory covariations or latency. Here, the cross-correlogram peaks simply demonstrate that the two spike signals are not independent and, therefore, do not correctly quantify the associations between the neurons. Moreover, this method presumes that the naturally varying properties of neurons are time-independent. Justification of such an assumption that makes the spike signals unvarying with respect to time is difficult since a time-altering stimulus is used to evoke the response from neurons which tend to

regularly adjust to this given stimulus. Although an attempt has been made to address the variation in the spike signals by estimating the covariance with the help moving windows, it would again require enormous quantities of data.

[12] measures association in nerve cells using two distinct data. The first involves the use of sound-based stimulus on cats to record data from 8 neurons and the second involves data taken from the nerve cells of sea hares. Cross-intensity is used to define the dependence between two discrete spike events (point processes) which is a quantification based on histograms. An advantage of this approach is the simplicity in the calculation of the function measuring the association between the spike trains in pairs. However, even though the Cross-intensity provides an estimate of the confidence interval between the spiking activities, the approach has received minimal attention in the neuronal analysis[12].

The peri-stimulus time histogram or simply PSTH and raster plots depict the number of spikes at all points of time. Preference towards the PSTH arises from the ability of our eyes to better articulate the plot showing the rate of firing of neurons. An extension to this concept is the joint-PSTH (See Fig **??** c) which jointly displays the number of spikes for a pair of neurons, the diagonal of which shows the simultaneously observed firing rate varies as per the specified width of the bins. [13] provisions methods for filtering the data from neurons using Bayesian adaptive regression that results in smoother plots and further gives a better estimate of the instantaneous degree of spiking.

In [14], a procedure to quantify the time-dependent neuronal spiking rate is introduced which explains a distinction between indirect and direct effect of the stimulus. The former is explained using the normalized Joint peri-stimulus time histogram (JSPTH) while the latter is depicted using a JSPTH predictor performing cross-product of the PSTHs. These depict that there is no significant difference between the 2 neuronal spiking activities. The normalized JSPTH is moderation to the JSPTH and its calculation is performed by subtracting the bias (caused when the neuronal activities are independent) from the joint rate of spiking. And a ratio of this result is taken over the multiplication of deviation of the 2 spiking activities. Such a correction is conducted keeping in the account for the elevation in the spiking rates (a sign of better associativity) when the spike trains are independent of each other. Pearson correlation is used to compute the normalized JPSTH for a different number of trials. [14] also

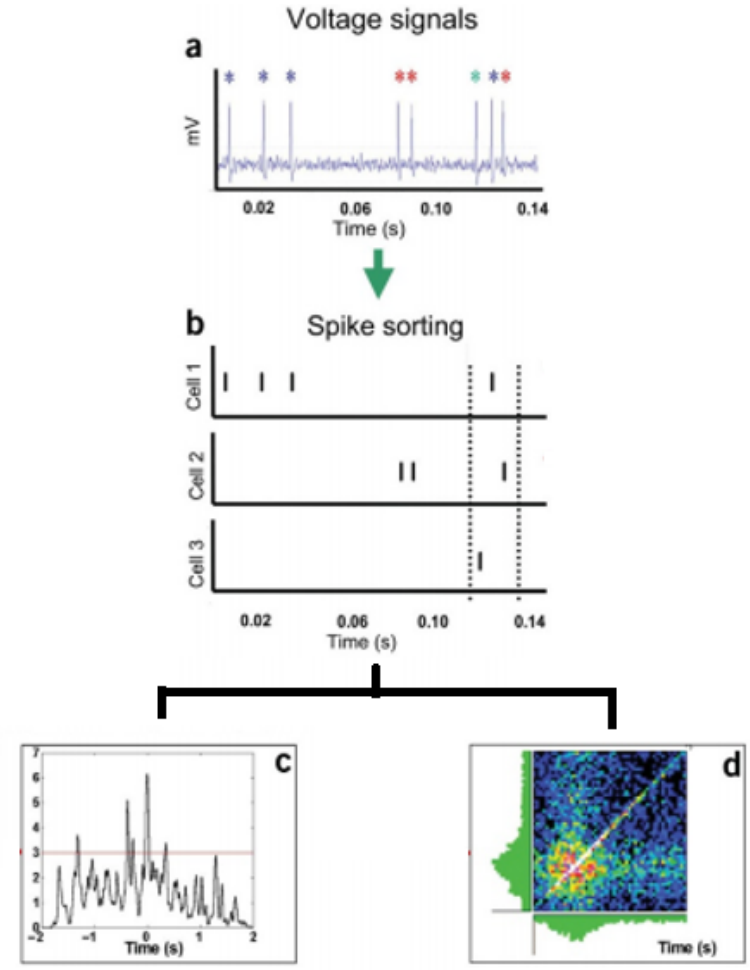

Figure 2.1: Measuring interactivity between neural signals using JSPTH and maximum likelihood based approaches.[1]

compares the normalized JSPTH with different normalization method and demonstrate that their approach outperforms them. The normalization is naturally independent of the model chosen and has symmetricity with correspondence to the spike train of neurons. Importantly, taking a sum of the diagonals of this measure creates a normalized covariogram. Furthermore, a study on the quantification of interdependence using JSPTH without involving the rate of spiking is conducted in [15]. Here, simulation of spike train data was used to describe that normalization of JPSTH wrongly explained the time-based structure in the strength disparities of the interaction parameter. The simulated data was produced using a model based on the probability of the Interaction

of discrete temporal events. Therefore [11] demonstrates that even though normalizing covariogram and JSPTH is beneficial, each of them has its shortcomings. This demonstrated that firing rate modulations cannot be corrected for in a model-independent manner. [1] highlights the limitations by primarily suggesting that the precision of measures relies upon the process generating joint firing activity and is one among the many measure that can be used for association. Besides, based on the presumptions and the underlying approaches, the methods of testing the statistical significance can vary. Also, the normalization of the covariogram and JSPTH are based on an assumption that the corresponding trials of spike trains cannot be distinguished statistically. Hence, if there is a change observed in the spiking patterns of the corresponding trials, then this change occurs as synchrony. Lastly and importantly, all analysis of interdependency of spike trains is conducted based on a prior sorting of the spikes, precision of which is crucial. However, most of the algorithms used for conducting spike sorting generate ingenuine correlations among the neuron pairs.

Moreover, several classification algorithms are used to find accurate patterns in the temporal spiking activity which in turn measure the associativity between the neural signals [16]. These approaches provision for measuring the interactivity between 2 or more neural interactions further allowing research related to behavioral happenings [1]. However, a challenge in using these methods is to contemplate the statistical test and the no-difference (null) hypothesis as well as decide on the pattern complexity, which is why findings in some studies based on these analyses have often received criticism. Apart from this, statistical modeling tools like likelihood (See Fig **??** d) have also been employed to study the interactions between neural point processes [17]. These approaches can be used for performing data analysis in the neural apparatus, given to the availability of a large variety of well-built statistical methods that assess how well a model fits, build confidence intervals and test hypothesis [1]. However, it is difficult to define multivariate discrete time-based models that precisely mu. Moreover, it is also challenging to define effective algorithms for fitting models.

2.2 Coherence and Synchrony

This section briefs on literature that makes use of linear measures like coherence and autospectrum to estimate the interdependency between spike train.

2.2.1 Coherence

Coherence is a frequency-domain measure that determines the relationship between two temporal events (as in the case of neural signals). By conducting a Fourier transformation on the neural spike data, an analysis based on the frequency of the combined firing activity of the neurons can be performed, which is further employed to calculate the spectrum of single spike signals or to compute the coherence of the pairs of spike trains [1]. This approach of measuring the associativity comes with 2 benefits the primary one being that normalization is independent of the width of the bins and pairs of spike trains can be combined offering pooling across the data. The second is that the different measures are not required to quantify the dependence of the process with discrete temporal events, or processes taking continuous values, or hybrid processes. Furthermore, [18] also describes that an estimation of the confidence intervals and error rates for cross-spectrum is valid only if the count of peaks is high.

[19] demonstrates that non-variation in the spike trains concerning the given stimulus is crucial to showcase the relationship between neurons and suggests making use estimations of coherence based on moving windows to analyze such non-stationarity. Furthermore, the principle of uncertainty on the coherence evaluations in a frequency-time domain is described for the estimations based on the windows that lay a lower limit on the spread of the point spread functions. Compared to the time-dependent estimations A lower bias is observed in the evaluations based on the moving window. This also suggests that by performing inverse Fourier transformation on the frequency-dependent method, optimal estimation of the time-based functions can be achieved.

Furthermore, [20] describes the use of multitaper scheme for the quantification of frequency-based measures. Such a principled method can also be utilized for the calculation of coherence between the neuronal spike data as well as for the estimation of potential in the local field. Moreover, with the help of bootstrap techniques, the method also proves useful in deriving the error rate of the estimations. explains this by studying the activities of spike train elevations within cortical regions while a memory task is performed. The repetitions in the spiking are investigated by analyzing area LIP and the time-based activity of potential in the local field. Important revelations in the study include the following: activities in the local field potential show that memory fields with dynamic nature have a spatial tuning in the gamma region which are not

present at lower-ranged frequencies. Instead of being fixed, spiking Activities of the neurons have a time-based structure while the memory task is performed. Decoding of potentials in the lower fields on individual trials was conducted and it was deduced that the activities in parietal cortex distinguished the chosen and the not-chosen direction with almost equal precisions as the rate of firing. This helped in producing a closer prediction of the planned movement. Moreover, for the same phase, coherency is observed between the lower field potential and firing activities of neurons in the gamma region.

An approach independent of the parameters and of the time scale called the ISI distance is described in [21], which is a complementary method that estimates the interdependence of the spike trains by calculating the ratio of the instantaneous spiking activity. This is demonstrated by using the method to recorded cortical neurons which makes the easies visualization of the firing pattern. The ISI distance is compared with other approaches and it is concluded that is more beneficial as none of the parameters dont need to be optimized. Indeed, the approach automatically adapts and spots the right time scale and also holds for the variations in the spiking patterns consistent firings and bursts which are incorrectly estimated by another measure whose behavior varies as per the selected parameter. However, the ISI-distance can be defrauded by the lags in the phase, not depending on whether the cause is latency loops in a single spike channel or if the same driver enforces varied delays on the two point-processes.

2.2.2 Power Spectral Density

A Power Spectral Density (PSD) determines how the power of a signal is distributed with respect to the frequency components. [2] performs power spectral analysis on two patients suffering from brain injuries with minimum consciousness and self-awareness. Such patients can evoke from their state of coma and therefore it is necessary is distinct the 2 states to provide for the treatment. EEG power spectrum was studied when the patients were sleeping and while they were to understand the behavior of the subconscious state. The frequency related part of the time-dependent EEG signals are summarized and a measure of the relative power of contributing frequencies against the signal composition is indexed. Multitaper approaches were employed to estimate the power on data taken for 3 seconds (see Fig **??**) and 5-second interval being sampled at

200Hz. These approaches apply numerous data tapers for the bias optimization and for stabilizing the variance of the spectral estimations. After performing tapering. an average over each spectral estimate is taken to calculate the power spectrum. This is given by

$$s_k(f) = \sum_{i=1}^{N} v_t(k) s_t \exp(-2\pi i f t) \tag{2.1}$$

With s_t represents the signals and v_t is the respective weights of the sequential tapers. The power spectrum is calculated by taking an average of the tapered estimates.

$$S_{MT}(f) = \frac{1}{K} \sum_{k=1}^{K} \left| s_k(f)^2 \right| \tag{2.2}$$

Observing the power spectrum peaks shows dynamic disturbances in the β frequency when the first patient is awake. The EEG based discoveries suggest a possible base for the dissociative lower metabolism which provides the suggestion of utilizing stimulus-based mechanisms to assess EEG spectra potentials that change with the state.

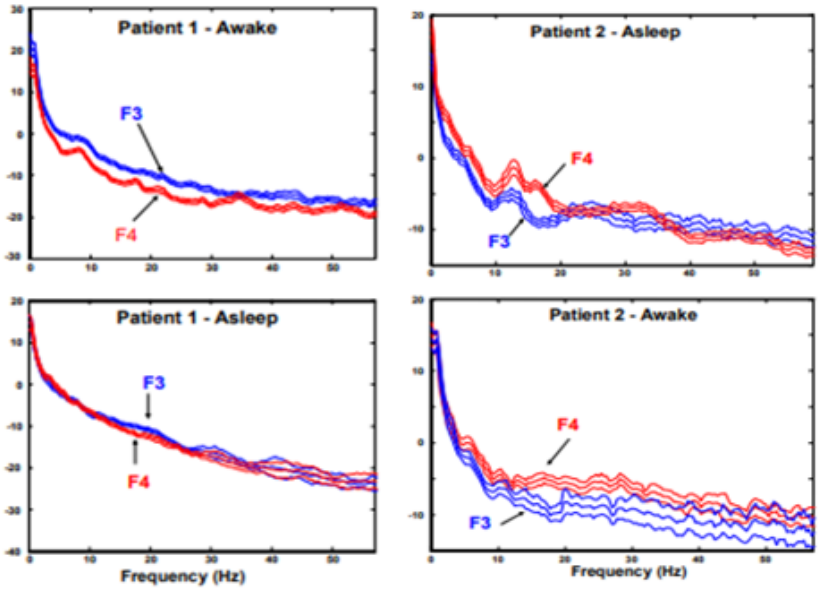

Figure 2.2: Power spectral desnity of Patient 1 and 2 during the states - awake and asleep for a swatch of 3 seconds [2]

[22] aims to showcase a pattern in the power spectral density of magnetoencephalography signals which reduces with aging. This allows answering how the resistance of skull decreases with development. By performing the Fourier transform of the method that autocorrelations every epoch of MEG, an estimation of the power spectral density were performed. Thereafter, an average of the power spectral density of every trial was taken into different frequency bands. This helped in determining whether the changes based on age in the higher and lower frequency ranges are genuine or are rather being created due to the statistical reasons (based on age-based shift). To lower the count of parameters in the analysis, mean power spectral density values within the different bands were divided into five groups based on the sensor location. [22] illustrates that power spectral density reduces with aging which can be due to the intensity decrease of the sources of neural signals being related to the pruning of synapse. Moreover, analyzing the correlations demonstrated that a late maturity in the rhythmic movements in the brain is seen in the frontal regions in comparison to the other regions.

2.3 Entropy and Mutual Information

Non-linear information theory measure like the entropy and mutual information are also used to measure the dependencies between neural spike signals. These measures have close correspondence to the neurological studies as given to electric signals recorded from nerve cells having an underlying non-linear structure. This section provides a review of the literature of these measures.

[23] used entropy to measure the synchrony between spike signals collected from neurons. The information generated of the time of spiking activity is quantified in bits which is independent of the presuppositions on the importance of features of spike signal. The data is recorded from neurons of the visual system of a fly. The study aims to provide estimations of entropy in neural signals that do not depend on the chosen model. Primarily, the dependency on the quantity of data is studied, thereafter, regularities in the behavior of the estimates are explored. This study is one of the pioneering studies that demonstrate the usage of non-linear information theory measure on the neural spike trains

[24] introduces two upper bounding mechanisms for the estimation of mutual information between neural spike trains. It is considered for some time T, neural response

R is measured with respect to the given stimulus S that takes continuous values. The primary approach, simple bounds calculates the stimulus from the response R. The method makes use of data processing inequality which a measure that holds if accurate upper bound is known or the mutual information between R and S is known. The secondary method applied is the Gaussian bounds that assume that if the stimulus takes Gaussian values, then a Gaussian response is observed with the noise being independent of the stimulus. The Gaussian bound Is more applicable to the neural spike trains that contain discrete values. Although mutual information evaluates the interdependency between S and R, a direct calculation of this measure is challenging since the precise estimation of the conditional probability, p(R—S) is required. This holds when responses contain values that are dependent on other neurons. The findings suggest that the estimation of mutual information using these methods depends on the estimator and can generate trivial estimates with bounds of 0. Also, these techniques make strong suppositions on the neural code suggesting that the estimates of mutual information are not reliable.

[25] performs a study on the two motor-output layers L2/3 and L5a to examine their neuronal activities in L2/3 and L5a while a task in learning. Calcium-imaging is performed on a mouse for a lever-pulling task. Lever movement is predicted utilizing support vector regression. Post this mutual information is calculated between the recorded and predicted lever trajectories with the help of bin-based method. It is observed that the accuracy of the predictions for L2/3 remained constant throughout the 14 trails, and some neurons depicted greater accuracy during data training. On the other hand, L5a showed a steady improvement in accuracy with 33 percent of its neurons significantly contributing to the overall prediction.

[26] showcases the use of measures for the estimation of mutual information that does not rely upon the presence of a coordinate system. This was defined as the density-based approach which is the method employed in this study. The author begins by providing measures for the estimation of entropy for 2 random vectors that represent neuronal spike trains. Thereafter an estimation of the volume is given, which was to be used for calculation of mutual information. Thereafter mutual information was defined for the vectors when either of them takes values in a coordinate system and when both take values in a coordinate system. Finally, an estimate of the bias using the Kullback-Leibler divergence is given for the two random spike-train representing

vectors.

[6] performs the calculation of mutual information using the density-based approach on fictitious spike train data. Values with closer proximity are found for each paired data value. The counts of these are used to perform the overall estimation of mutual information.[8] also provides a method to calculate the bias in the estimations. Further, a comparison of MI results using the two distance metrics - the Victor Purpura and the von Rossum is shown.

2.4 Neuro-electric signals

Several electrodiagnostic studies that record electrical activities from nerve cells such as the nerve conduction studies, blink reflexes, electromyography (EMG), electroencephalography (EEG) and others are conducted on patients suffering from neurodegenerative diseases. Among these researches, EMG forms the core of those related to neuromuscular disorders [4].

2.4.1 Electromyography

Electromyography(EMG) is a neurophysiological tool used to examine the electrical activity generated by the skeletal muscles(see Fig. **??**). An electromyograph is used to collect the EMG which detects the electric potential from the electrically or neurologically activated muscle cells [27]. The signals captured by the EMG are useful in detecting a medical abnormality, levels of activation, and are used to analyze the biomechanics of human or even animal movement.

The typical operating procedure for EMG makes use of electrodes which are placed on the skin above the muscle(see Fig. 2.4). These electrodes detect the electrical activity in a muscle which gets displayed on the monitor screen. The EMG also consists of preamplifiers and amplifiers which are placed in close proximity to the patients to avoid capture of any electrical interference and are used for achieving with correct gain. Post amplification, the signals are calibrated, after which it can be visualized on display screen, Other equipment consists of integrators that comprise the data obtained and a recording medium which is a fiber-optic photographic system.

The muscles are isolated leading to each electrode giving only an average knowledge

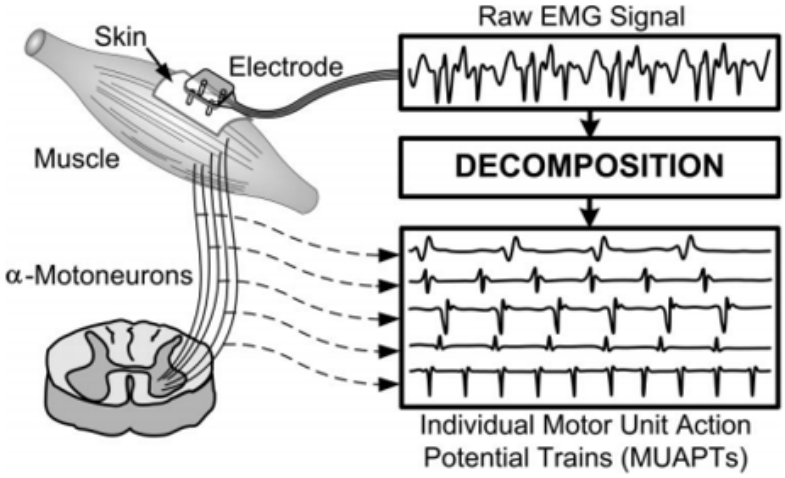

Figure 2.3: A typical EMG recording [3]

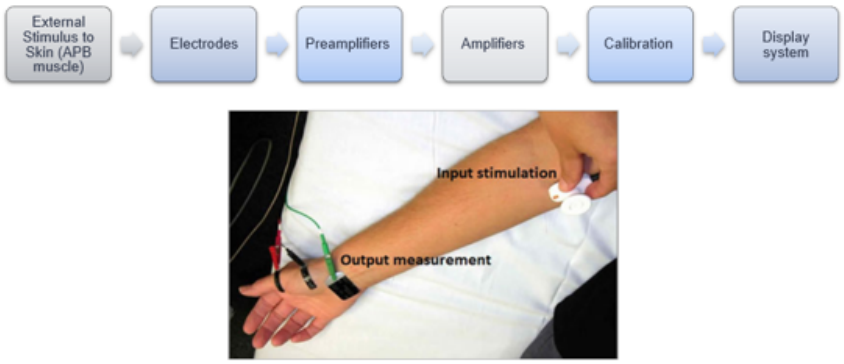

Figure 2.4: working of EMG [4]

of the activity of the muscle selected. Thus, there is a need for several electrodes to be placed at various specific locations to obtain accurate activity of the muscle. After the placement of the electrodes, the patient is asked to contract the muscle under the study. The presence of the waveform, the size of the waveform and the shape of the waveform gives specific knowledge about the muscles ability to respond to stimulation from the nerve. When the muscle is at rest no action potential is generated but when the muscle is externally stimulated there is a generation of the action potential which is observed at the display screen [28]

Types of EMG

EMG devices record the electrophysiological activity of the muscle which is innervated by the axonal branches of the motor neuron. Two types of electrodes are used to recording one is needle electrode and other is a non-invasive surface electrode.

- Needle EMG

  Needle EMG allows the recording of deep muscles with the help of insertion of a needle electrode into the muscle tissue. Location of the point of needle insertion is determined through the identification of anatomic landmarks that are further confirmed through contraction of the selected muscle to be measured. It has been observed that with needle EMG, more sensitive and accurate data is generated. Also, with its different types of muscles can be recorded at the same time [29]. However, needle EMG has several drawbacks, firstly it pertains to the activity of only a small number of muscles that have their fibers close to the position of electrode [30]. Secondly, needle EMG is a painful procedure and so prolonged recording with needle EMG is not convenient. Rarely, there have been changes in local trauma that have happened while examining delicate regions. Needle EMG is sensitive to time and temperature. The signal detected in the needle EMG will vary with respect to the time elapsed from the onset of the nerve injury. Temperature can affect data transmission within the muscles and nerves which in turn affects the action potential being carried through the muscle [31].

- Surface EMG

  Surface EMG (see Fig. 2.4)is a technique that is used to measure the activity of the muscle noninvasively with the help of surface electrodes. Surface electrodes are placed on the skin and over the muscle for which electrical activity needs to be detected. This has several benefits, primarily, this technique is painless, especially when used without the stimulation of peripheral nerve stimulation. Second, the electrode of surface EMG records from a wide muscle area giving global data of the functioning of the muscle. And finally, with surface EMG prolonged recording from multiple sites is possible. However, surface EMG has a low resolution of the signal, is also susceptible to movement artifacts and the temperature of the body. Also, with surface EMG it is possible to measure the

superficial activity of the muscles and so, the deeper muscles are not recorded. Conditions like obesity and edema cause an increase in the skin thickness leading to a disturbance in the recordings [32].

2.4.2 Application of EMG in neuroscience and neurological diagnosis

The muscle membrane potential is the source of the signal that arises in the EMG. EMG enables good quality temporal resolution and has an advantage in diagnosing central nervous system (CNS) that provide signals to muscles. The lagging in the signaling back from the CNS to the muscle is detected thus, aiding diagnosis of movement disorders like tremors and dystonia. complex movements recorded by the EMG helps in detection of gait disorders. Disorders like orthostatic tremor are also possible to diagnose with EMG but difficulties can arise because the frequency of shaking is 16 high hertz, and doctors cannot differentiate the tremor as the output generated is in a fused form. Therefore, EMG is clubbed with accelerometer leading to a clear presentation of tremor [33].

Advancements were required in the electrodes that were utilized to measure the defects related to the peripheral nervous systems signal to the muscle. This lead to the involvement of mono-bipolar EMG to detect the contractions in the muscle through choice rather than being involuntary. Carpal tunnel syndrome, ALS, Myasthenia gravis, dermatomyositis, Duchenne's muscular dystrophy, Guillain-Barre syndrome, peripheral neuropathy, Shy-Drager syndrome are among other disorders that can be detected through EMG in an asymptomatic stage [34].

2.4.3 Utility of Spectral frequency-dependent measures with EMG

EMG signals are often used to study the motor control of patients suffering from a genetic muscular disorder and other dystrophies like stroke. This section provides a brief discussion on such a use of the frequency domain measures - Coherence and Power Spectral Density of EMG signals.

In [35], a study on the EMG data of 2 regions of index fingers was conducted

while the muscles were contracted under fatiguing conditions. Coherence among the finger flexor muscles and the flexor digitorum superficialis EMGs were examined for all subjects in the respective gamma, beta, and tremor frequency regions. No important differences in the coherence were observed in the tremor regions and post contraction, a rise incoherence in gamma and beta frequency regions had been seen. An increase in Force-EMG cross-covariance in the tremor regions during and after the contractions were observed which proposed that corticomotoneuronal drive held the responsibility to the rise in the coherence in the other two regions. Moreover, an increase cross-covariance suggested that surrounding afferents contributed more to the fatiguing muscle coupling.

[36] aims to compare the EMG power spectrum to the physical measurements taken from several features of muscle potential in 18 healthy subjects and 32 patients, those suffering from Duchenne muscular dystrophy (which is a genetic muscular disease). This would enable the assessment of the diagnostic output from both the approaches. Needle EMG was used to collect data from the tibialis and biceps at weak contraction. Power spectral analysis using EMG signals was performed using a Haning window of 1400 Hz, while simultaneously, a visual analysis of twenty potentials by muscle units was recorded. In contrast to the healthy subjects, the EMG from the patients demonstrated lesser averaged amplitude of signal while in the case of physical measurements, higher potential having smaller duration could be seen. About three-quarters of the patients were correctly classified as myopathic using the physical-measurement approach. Power spectral analysis displayed a significant fall in total power while a significant rise could be seen in the relative and average power frequently, which classified a higher percentage 95% of patients as myopathic. Thus, power spectra of EMG yielded higher diagnostic output than by physically measuring the muscular potentials frequency. Furthermore, with respect to the control signal, the power spectra of EMG moved to greater frequencies in the patients.

By performing a power spectral analysis on the surface EMGs of stroke-induced patients, [37] aims to study the complexity in muscular and neural variations. The research was performed on fourteen subjects and EMG data were recorded from first dorsal interosseous and paretic muscle on voluntary contraction. To identify characteristic traits, power spectral analysis was conducted on the EMGs of FDI and paretic muscles. A distinction in the distribution patterns of the spectrum could be visualized

between the 2 muscles. While no significant spectral evidence could be found in some subjects, nine of them demonstrated a rise in the average power frequency in FDI muscle. Summing up the results showed a decrease in the average power frequency in paretic muscle compare to the FDI muscle. [37] suggests that patients affected by a stroke demonstrated neural and muscular processes produced an impact on the power spectrum. Also, at the same levels of contraction, almost all the subjects showcased more average power frequency in the FDI muscle than paretic muscle, which can be due to various reasons like degeneration of the muscle fibers, impairment of bigger units, disruption in the muscle control features and, higher synchronization between the muscle units. Such results depict EMG as a convenient tool to demonstrate stroke-related variations in the muscles and neural regions.

2.4.4 Application of Non-linear Information Theoretic measures with EMG

Among several other applications, EMG signals have also been widely used in the study of prosthetic devices for limbs, hands, and arms. This section briefly discusses studies on the use of information theory metrics like entropy and mutual information in medical diagnosis and prosthesis research work. Examining the activities of muscles in the respiratory system help to effectively research lung-related diseases like Obstructive sleep apnea syndrome (OSAS). Given the complex nature of the respiratory system, dysfunctions concerning OSAS are insufficiently assessed by linear approaches. Hence, [38] aims to detect complex respiratory processes with nonlinear characteristics having relevance to diagnosis. While performing an increased respiratory action, EMG signals of 8 healthy subjects and patients were recorded from 3 muscles of the respiratory apparatus. The objective is achieved by first, assessing the coordination of contractions in muscles within the respiratory region through nonlinear examination by estimating the cross mutual information and second, by differentiating the working of muscles in healthy subjects from the patients suffering from OSAS. The study involved several parameters and estimated the cross mutual information to define the dependency between EMG signals. Findings depict various nonlinear couplings in both the categories of subjects which become more visible when the respiratory effort is even further increased.

With EMG signals being widely used a control signal in prosthesis performing multiple functions, a challenge lies in them being able to precisely handle single and combined movements of the fingers, that too in a computationally optimal way. To address this challenge, an algorithm called Mutual Component Analysis is introduced in [39] that performs both parameter selection and result projection and is an extension to the approach Principal Component Analysis. EMG data were collected from the forearm of 8 patients was used to analyze the significance and efficiency of the algorithm. Redundant and less important parameters along with the noisy data were pruned, post prediction was performed. Mutual information was used for estimating the information gain, a quantity which was further used in the selection of parameter. [39] concludes that the strength of the algorithm could be witnessed in it achieving less than 5% error rate on 15 types of finger movements for all the 8 subjects.

Chapter 3

Methods

The section primarily provides measures for the estimation of baselines - Correlation, Power Spectral Density, and Coherence. This is followed by a discussion on the two approaches used in this study: the bin-based method using the Akaike Information Criteria and the Density-based approach using Kozachencko-Leonenko Estimators.

3.1 Baseline Estimation

Information theory provides several measures to estimate the interdependency between EMG channels. This section briefly explores them as follows

3.1.1 Correlation

Correlation defines a measure to estimate the linear relationship between 2 random entities. In the context of neuroscience, this metric measures the functional dependency between electrophysiological data like electroencephalogram and electromyography [40]. The EMG is a representation of continuous values of voltage per unit time and is treated as multivariate time series as well as it belongs to random processes which can be better described by probability distributions rather than deterministic mathematics. EMG enables for the estimation of correlation as well as coherence allowing successive points to be independent in time and this time-dependent random process can also be extended to the spatial domain.

To estimate the correlation between spike trains (at a particular frequency), the signal needs to be filtered by specifying the count of frequency bins or bands, either by performing digital filtering or by using analog devices. If analog devices are used either of the two requirements needs to be satisfied: first, every frequency should have its distinct analog filters or based on the required bands, second, the signals must be repeated multiple times.

For a specified frequency s, the function for correlation is defined as:

$$g(s) = \frac{C_{PQ}(s)}{(C_{PP})\,(C_{QQ})} \tag{3.1}$$

where C_{PP} refers to the auto-covariance of the spike channel Q; C_{QQ} refers to the auto-covariance of the spike channel Q, and C_{PQ} refers to the cross-covariance of the spike channels P and Q

The autocovariance (which gives the joint variability of a signal with itself) for the respective spike trains P with mean $\mu_P(e)$ with time moments e_1, e_2 being the time moments is given by [41] as

$$\operatorname{cov}\left[P_{e_1} P_{e_2}\right] = E\left[P_{e_1} P_{e_2}\right] - \mu_P\left(e_1\right)\mu_P\left(e_2\right) \tag{3.2}$$

where E is the expectation operator. Similarly, the autocovariance for Q with mean $\mu_Q(e)$ is

$$\operatorname{cov}\left[Q_{e_1} Q_{e_2}\right] = E\left[Q_{e_1} Q_{e_2}\right] - \mu_Q\left(e_1\right)\mu_Q\left(e_2\right) \tag{3.3}$$

The cross-covariance provides temporal variation of one signal against the other [41]. The cross-covariance between spike trains P and Q is given by

$$\operatorname{cov}\left[P_{e_1} Q_{e_2}\right] = E\left[P_{e_1} Q_{e_2}\right] - \mu_P\left(e_1\right)\mu_Q\left(e_2\right) \tag{3.4}$$

Correlation is independent of the measure of amplitude and varies with respect to both polarity and phase. With a reduction in polarity information, the estimation of correlation produces values ranging from -1 to 1.

3.1.2 Power Spectral Density

The power spectrum or commonly known as the auto spectrum determines how the power of a signal is distributed with respect to the frequency components [42]. As per Fourier analysis, a continuous signal can be fragmented into a spectrum of frequencies. The spectral resolution is used to normalize the amplitude of power spectral density and is used to distinguish random signals.

PSD is applied over a signal over its entire period measurement (which is usually large), which then accounts for the distribution of spectral energy per unit time, an integral or sum of this provides cumulative power [43]

For L_t point processes, PSD X_{tt} for EMG spike train at frequency s [43] is given by

$$X_{tt}(s) = \frac{G_1}{2\pi} + \frac{1}{2\pi} \int_{i=-\infty}^{i=\infty} r_{tt}(v) e^{-isv} \tag{3.5}$$

where G_1 represents the mean intensity while $r_{tt}(v)$ represents the autovariance of a given spike channel.

3.1.3 Coherence

Another linear measure that defines the dependency between 2 spike trains is the coherence. With the advent of faster computational algorithms like the Fast Fourier transform, methods like correlation were succeeded by an alternate statistical method Coherence [40]. The advantage of the method lies in it depicting the covariation between 2 spike trains in a shorter period producing results similar to those produced by correlation. Its several applications lie in the area of clinical diagnosis, in understanding disorders in psychiatric patients, studying cognitive function, and for understanding intermuscular relationships (between EMG channels)

For a specified frequency s, the function for coherence (h) is defined as:

$$h(s) = \frac{|X_{PQ}(s)|^2}{(X_{PP})\,(X_{QQ})} \tag{3.6}$$

where X_{PP} is the auto spectral density of the spike channels P; X_{QQ} is the auto spectral density of the spike channels Q and X_{PQ} is the cross-spectral density of the spike channels P and Q

$$X_{PQ}(s) = \frac{1}{2\pi} \int_{i=-\infty}^{i=\infty} r_{PQ}(v) e^{-isv} \tag{3.7}$$

where $\mathrm{r_{PQ}}(v)$ is the cross-covariance of the spike channels P and Q

Coherence produces results that closely correspond to correlation. However, coherence is dependent on phase relationships and variation in power since it is estimated by the ratio of the square of the cross-spectral density to that of the two auto spectral densities [40]. This suggests that the value of coherence is sensitive either to phase or power variation, a sudden asymmetry in which is not anticipated under regular physiological conditions. Apart from this, for a particular epoch, the coherence always remains one. Although being dependent upon phase and power of paired spike signals in consecutive epochs, if no change is recorded over time, the value of coherence does not vary and remains one. This suggests that coherence lacks in providing exact information on the direct relation between two spike trains, only ensuring the stability of the relationship concerning the phase and power metrics.

3.1.4 Mutual Information

Information theory suggests that mutual information between 2 random variables measures the relationship between them. This holds relevance to neuroscience as mutual information provides can be useful to examine electrophysiological data such as spike trains from neurons (which can either be from a single one or from multiple). Although many statistical metrics define the dependence between the spike trains like the rho by Spearman, the correlation coefficient by Pearson, and Kendalls tau, mutual information is more practical since it doesnt require any presumption based on marginal distributions [44, 25]. Also, being an additive metric, the average of this measure makes more sense.

Using the joint density function *p(a,b)* along with the marginal density function *d(a)*, *q(b)*, the mutual information between a and b is defined as

$$I(a,b) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(a,b) \log \frac{p(a,b)}{d(a)q(b)} \tag{3.8}$$

Fig. **??** shows a comparison of the linear versus non-linear information theory measures: mutual information and correlation. Model 1 shows that when the two

random variables have a linear relationship, the correlation and the mutual information show sufficiently finer and coherent results. However, when the interactions between the data become non-linear (as seen in models 2 and 3), the linear measure tends to show inaccurate results despite the presence of underlying relationships in the data. On the contrary mutual information captures these variations and is, therefore, a more reliable measure for the estimation of interdependency.

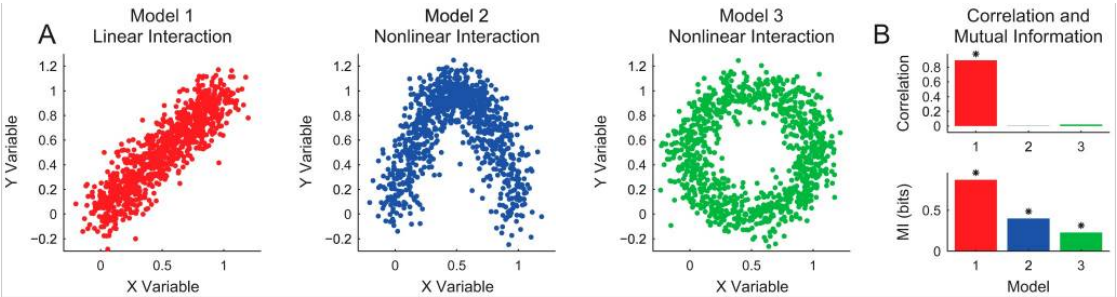

Figure 3.1: Linear versus Non-Linear information theory measures [5]

3.2 Bin-based Approach (using Akaike Information Criteria) for calculating mutual information

For the calculation of mutual information, the bin-based method performs discretization on the sample. Spike trains are converted into small bins of time, t over windows of length T, hence transforming the entire neural response into words [45]. The count of spikes concerning each bin is calculated and an estimate of mutual information is determined using the probability of how often a word reoccurs in data. This Bin-based approach uses a confusion matrix that keeps the counts of binned spike train data against the discrete stimulus entities.

In Bin-based methods, calculation of mutual information heavily depends on the discretization of the data distributions. However, within a small spike train interval, the discretized distribution of EMG data has dense values which suggest that bias in calculating the mutual information would be high. Hence, to discretize the continuous EMG data such that it does depend on the marginal distribution shape, the copula method is used. Eventually, to provision for discretization, the size of bins is

determined. Subsequently, the calculation of mutual information is performed.

3.2.1 Copula Estimate

Copula C was proposed to show the relationship between two random variables and are simply marginal distributions with multiple variables that have uniform one-dimensional margins bounded between (0,1)[46]. Copulas are extremely useful in applications (statistical) dealing with higher dimensions wherein by calculating marginals, these can be used for estimation of the distribution of random vectors.

Consider two random variables A and B with respective marginal distribution functions $G(a) = P(A \leq a) = r$ and $\Gamma(a) = P(B \leq b) = s$ and, and joint distribution function $D(a, b) = P(A \leq a, B \leq b)$. Here, every pair (a, b) corresponds to the three functions $G(a)$, $\Gamma(b)$, and $D(a, b)$ whose values lie between 0 and 1. With respect to $D(a, b)$, for every pair (a, b), there lies a point $(G(a), \Gamma(b))$ in a square of unit area with dimensions $[0, 1]X[0, 1]$. These joint distribution functions and marginals distribution functions are coupled by the copula method [25] as follows:

$$D(a, b) = K(r, s) = K(G(a), \Gamma(b)) \tag{3.9}$$

Moreover, copula density $K(r, s) = \frac{\partial^2 k(r,s)}{\partial r \partial s}$ links marginal density functions to the joint density function as follows,

$$p(a, b) = k(r, s)g(a)\gamma(b) \tag{3.10}$$

With the help of copula method, the mutual information [25] is given by

$$\begin{aligned} I(a, b) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k(r, s)g(a)h(b) \log k(r, s) dadb \\ &= \int_0^1 \int_0^1 k(r, s) \log k(r, s) drds \\ &= -H(k) \end{aligned} \tag{3.11}$$

where $H(k) = -\int_0^1 \int_0^1 k(r, s) \log k(r, s)$ defines the copula entropy. Therefore, by calculating the copula entropy, one can estimate the dependency between 2 continuous

variables using mutual information.

Given a w sized sample represented by $\{(a_l, b_l)\}_{l=1}^{w}$ be taken from a continuous bivariate distribution, the empirical copula frequency, k_w is

$$k_w\left(\frac{e}{w}, \frac{f}{w}\right) = \begin{cases} \frac{1}{w}, \text{if } (a_e, b_f) \text{ belongs to sample 0} \\ 0, \text{ otherwise} \end{cases} \tag{3.12}$$

where $a_(e)$ and $b_(f)$, and $1 \leq e, f \leq w$, represent sample based order statistics.

To calculate copula density, primarily, binning of the empirical copula frequency is performed, which divides copula frequency into a histogram. Therefore, if the count of bins bounded by the interval $[0, 1]$ is m, then, The cumulative bins within the space $[0, 1]\times[0, 1]$ is m^2 and the area corresponding to each bin is $1/m^2$, thereby the estimation of the copula density [25], $k_m(r, s)$ is given by

$$k_m(r, s) = m^2 \sum_{e=\left\lfloor \frac{w(x-1)}{m} \right\rfloor}^{\left\lfloor \frac{wx}{m} \right\rfloor} \sum_{f=\left\lfloor \frac{w(y-1)}{m} \right\rfloor}^{\left\lfloor \frac{wy}{m} \right\rfloor} k_w\left(\frac{e}{w}, \frac{f}{w}\right) \tag{3.13}$$

Where the respective ceil and floor function for some t are represented by $\lceil t \rceil$ and $\lfloor t \rfloor$. Also here, $x = \lceil mr \rceil$ and $y = \lceil ms \rceil$

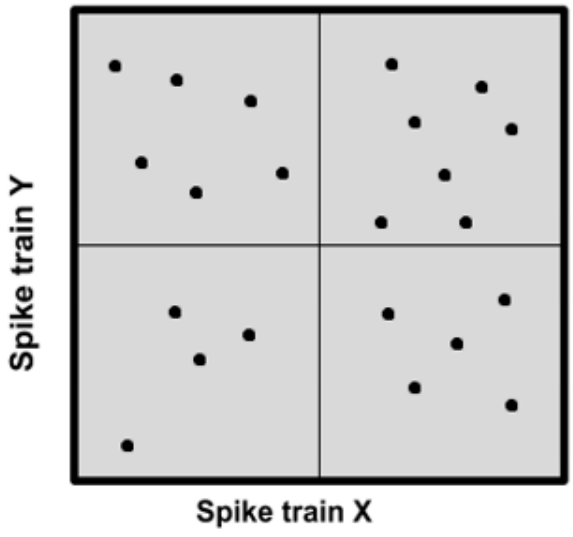

Figure 3.2: Cartoon explaining the working of Bin-based method

The above cartoon describes the bin-based method when the number of bins is 4. It can be visualized from the figure that each data point in EMG spike trains is associated with either of the 4 bins. Calculation of Copula density closely associates the data values to their respective bins. Summing up the data points with respect to the bins provides an estimate for the Akaike Information Criteria which is further used the estimation of mutual information.

3.2.2 Akaike Information Criteria

To obtain a fine estimate of the copula density, it is important to determine a fitting value of m, which relies on the selection of the model. Therefore, an estimation of the Akaike information criterion (AIC) for k_m is performed. Information is lost when models are used to express the data generation process. Hence, AIC provides an estimate of the model quality by providing the relative quantity of information lost by keeping a balance between how simple a model (underfitting) is, versus how better the model fits (overfitting) [47].

Although an estimate of AIC can be calculated with the help of some unknown variables by using the summation of the log-likelihood (negative), it is well established if the count of non-empty bins accounts for the unknown variables. The count of data points corresponding to a $bin(e, f)$ is given by

$$\frac{w}{m^2} k_m \left(\frac{e}{m}, \frac{f}{m} \right) \tag{3.14}$$

Hence, the likelihood of k_m is

$$L\left(k_m\right) = \prod_{\theta=1}^{m} \prod_{f=1}^{m} \left(k_m \left(\frac{e}{m}, \frac{f}{m} \right) \right)^{\frac{w}{m^2} k_m \left(\frac{e}{m} \frac{f}{m} \right)} \tag{3.15}$$

Therefore, AIC is represented [25] as

$$\mathrm{AIC}(m) = -2 \frac{n}{m^2} \sum_{e=1}^{m} \sum_{f=1}^{m} k_m \left(\frac{e}{m}, \frac{f}{m} \right) \log k_m \left(\frac{e}{m}, \frac{f}{m} \right) + 2J \tag{3.16}$$

Lastly, taking a negative of the calculated copula entropy, $H(k_m)$ provides an estimate of mutual information [25], I given by,

$$I(a,b) = -H\left(k_m\right) = \frac{1}{m^2}\sum_{e=1}^{m}\sum_{f=1}^{m} k_m\left(\frac{e}{m},\frac{f}{m}\right)\log k_m\left(\frac{e}{m},\frac{f}{m}\right) \tag{3.17}$$

3.3 Density Based Method

EMG data collected from a neuron or group of neurons holds values that lie in a metric space. As already discussed, for the calculation of mutual information, the conventional bin-based method performs discretization on the sample by converting spike trains into bins and transforming the neural response into words[45]. As the bin width, vanishes (with infinite data), the MI reaches its true value. Although such measures have been proved useful, estimating mutual information with such methods is difficult as enormous of data is required to estimate them. For instance, with 4s of EMG spike train data having a bin-span of 10 ms, there are between 2^{400} words and 2^{800} sets of words corresponding to the two spike-train interval pairs. With such a large no of words produced, projecting the probabilities of the entities would require amounts of data that would be practically infeasible to record.

Apart from this, information theory is mostly associated with problems where the data is mostly discrete. To address this difficulty, Kozachenko and Leonenko estimator is put to use which is a non-parametric differential entropy estimator that relies on n nearest neighbors (or commonly known as K Nearest Neighbour). Although the difficulty of estimating mutual information with discrete point processes has already been addressed [26], this approach provisions an estimator for an effective local dimension in line with the difficulty of requiring larger amounts of data. The efficacy of Kozachenko and Leonenko reflects in the fact that, for a metric space, this approach exploits the proximity structure, hence, considering the data points in pairs. This would allow for the calculation of mutual information of data like EMG that lie in metric space with a notion of similarity measure suggesting that the data would not have any coordinates but the distance between 2 data points could still be measured. Also, this approach produces results that are equivalent to the conventional bin-based method and takes a considerably lesser amount of data.

The density-based approach takes into consideration simple formulas[26] for the estimation of mutual information between

1. Two stochastic entities with one taking values in metric space and the other on discrete space.

2. And, where both the entities take values in metric space

To arrive at these formulas, primarily, a volume-dependent estimation of probabilities is defined using the Kozachenko and Leonenko method.

3.3.1 Calculation of Entropy using quantities with no coordinates

Consider a random variable A with a set of S outcomes, $a_1, a_2, , a_S$ having values in space A. If the probability mass density of this variable is given by $p_A(a)$, then the entropy can be estimated by

$$H(A) \approx -\frac{1}{S}\sum_{i=1}^{S}\log_2 p_A\left(a_i\right) \tag{3.18}$$

Since $p_A(a_i)$ is a quantity that remains unknown, the approach for the calculation of entropy is modified such that it is coordinate independent. A region $R(a_i, Vol)$ is taken in a manner such that such a region exists around every data point. Hence, the probability that this region[26] contains g data points is

$$P_g\left(r_i\right) = \left(\begin{array}{c} s \\ g \end{array}\right) E_i^g\left(1-E_i\right)^{s-g} \tag{3.19}$$

where $R(a_i, Vol)$ contains the probability mass, E_i.

This suggests that $\langle g\rangle = SE_i$, which can be calculated from the data as

$$\langle g\rangle \approx \#R\left(a_i, Vol\right) \tag{3.20}$$

where $\#R\left(a_i, Vol\right)$ defines the count of data points in $R(a_i, Vol)$

By assuming that the region (ball) has a constant probability mass function, $E_i = p_A(a)Vol$, upon further simplification, an estimate of the entropy [26] can be defined as

$$H(A) \approx \log_2 S + \log_2(Vol) - \frac{1}{S}\sum_{i=1}^{S} \log_2 \#\,(a_i, Vol) \tag{3.21}$$

The simplicity in the entropy estimation lies in the fact that instead of using the size of the region (containing a specified number of data points), the number of points in the regions is considered.

3.3.2 Calculation of Volume

The estimation of the volume of a region is determined with the help of probability distribution. This comes from the infeasibility in using the coordinate-based method since spike trains lack good coordinates. Thus, the volume of a region is given by the probability mass contained in that region,

$$\mathrm{vol}\, R = P(a \in R) \tag{3.22}$$

which can be calculated from the data as

$$\mathrm{vol}\, R = \frac{\#R}{S} \tag{3.23}$$

If $v_{ol} = h/s$ for some $h \leq S$, then an incidental estimate of entropy [26] can be determined as,

$$H(A) \approx \log_2 S + \log_2 \frac{h}{S} - \frac{1}{S}\sum_{i=1}^{S} \log_2 h = 0 \tag{3.24}$$

Since entropy is a quantity that relies on the measure, its value differs if the measure is changed. On the contrary, using probabilities for calculating information measure quantities is more practical if mutual information is used since its values dont depend on the measure used.

3.3.3 Calculating mutual information between 2 random variables where one takes values in a discrete space

EMG is essentially collected by either using an external stimulus on the hand or by recording a voluntary muscle contraction. In the former case, a discrete random variable can be used to represent the stimuli, while on the other hand, the response can hold values in a metric space. Hence, consider the recorded spike trains (either from a single neuron or from multiple neurons) have stimuli be depicted using a discrete set C and the responses using a set D. Also, let C have it's every unit represented the same number of times, n_t and let the count of stimuli be n_c. This accounts for the equivalent data points to be $S = n_t * n_c$.

The regions around the data points are defined in the response set, D. Thus, an open ball [26] is defined for a point d in D, which is given by

$$R_\epsilon(d) = \{t \in \mathrm{R} : \mathrm{dist(d, t)} < \epsilon\} \tag{3.25}$$

Also, consider that with a selected ε,$R_\varepsilon(d)$ has a volume Vol equivalent to the region $R(d, Vol)$. Using cumulative probability $p_D(d)$ as measure and fixing Vol as $\text{Vol } = h/s$ for some $h \leq s$. This suggests that h points lie in the ball $R\left(d, \frac{h}{s}\right)$ and with this, $H(D) = 0$.

Hence, the calculation of $H(C|D = d)$ is given by,

$$H(C|D = d) \approx -\log_2 n_c + \log_2 h - \frac{1}{n_t} \sum_{i=1}^{S} \log_2 \# \left[R\left(d_i, \frac{h}{S}\right) \right] \tag{3.26}$$

Averaging over $c \in \mathcal{C}$, the mutual information [26] between C and D is

$$I(C; D) \approx \log_2 n_c - \log_2 h + \frac{1}{S} \sum_{i=1}^{S} \log_2 \# \left[R\left(d_i, \frac{h}{S}\right) \right] \approx \frac{1}{S} \sum_{i=1}^{S} \log_2 \frac{n_c \# \left[R\left(d_i, h/S\right) \right]}{h} \tag{3.27}$$

The calculation is straightforward and can be easily put forth in the case where both the random variables hold values in metric spaces.

3.3.4 Calculating mutual information between 2 random variables both taking values in a metric space

The volume measures are given by the marginal probability mass function $p_C(c)$ and $p_D(d)$ such with which $H(D) = H(C) = 0$. This also implies a measure on the space $C \times D$, where the values (c, d) lie. Thus regions around (c_i, d_i) are defined using a square to determine p_C,$D(c_i, d_i)$ and the volume of the regions is estimated using $p_C(c)$ $p_D(d)$ which are taken from the marginal space of C and D. Hence,

$$C\left(c_i, d_i, \frac{h_1}{S}, \frac{h_2}{S}\right) = \left\{(c,d) \in \mathcal{C} \times D : c \in R_c\left(c_i, \frac{h_1}{S},\right), d \in R_D\left(d_i, \frac{h_2}{S},\right)\right\} \quad (3.28)$$

Where $h_{1/S}$ and $h_{2/S}$ are the selected volumes for C and D. So, with the implied measure,

$$\text{volume } C\left(c_i, d_i, \frac{h_1}{S}, \frac{h_2}{S}\right) = \text{ volume } R_C\left(c_i, \frac{h_1}{S},\right) \text{ volume } R_D\left(d_i, \frac{h_2}{S},\right) \approx \frac{h_1 h_2}{S} \quad (3.29)$$

Therefore,

$$I(C;D) \approx \frac{1}{S}\sum_{i=1}^{S} \log_2 \frac{S\#\left[C\left(c_i, d_i, \frac{h_1}{S}, \frac{h_2}{S}\right)\right]}{h_1 h_2} \quad (3.30)$$

This suggests that the mutual information between C and D relies on $S\#\left[C\left(c_i, d_{i,}, \frac{h_1}{s}, \frac{h_2}{s}\right)\right]$, which is the count of response-stimulus pairs (c, d), where c and d are one among the $h1$ points nearest to c_i and $h2$ points nearest to d_i respectively [26]. The values of the h_1 and h_2 should be chosen carefully as taking either small or large value results in the reduction of the accuracy of mutual information.

3.3.5 Kullback-Leibler Divergence

The KL divergence provides a measure of the difference between 2 probability distributions. The KL divergence for 2 random variables C and D on the same metric space that take values $c_1, c_2, \ldots, c_U$ and $d_1, d_2, \ldots, d_S$ is defined [26] by

$$d(C|D) \approx \frac{1}{U} \sum_{i=1}^{U} \log_2 \frac{p_C(c_i)}{p_D(d_i)} \tag{3.31}$$

Since,

$$UVolp_D(d_i) \approx \#[B(d_i, Vol)] \tag{3.32}$$

However, in the situation, the other distribution is used for calculating the volume. If h/s is taken as the volume, then the ball $\mathrm{R}[\mathrm{B}(d_i, h/\mathrm{s})]$ around the data point d_i is sufficient to include the h points from $c_1, c_2, \ldots, c_U$ and ${}_R[B(d_i, h_S])$ is the count of datapoints in the open ball [26]. Thus,

$$d(C|D) \approx \frac{1}{U} \sum_{i=1}^{U} \log_2 \frac{S\#R[B(d_i, h/S)]}{Uh} \tag{3.33}$$

3.3.6 Estimation of Mutual Information between paired-spike channels

Let $P = \{(\mathrm{m}_1, \mathrm{n}_1), (\mathrm{m}_2, \mathrm{n}_2), \ldots . (\mathrm{m}_\mathrm{n}, \mathrm{n}_n)\}$ denote the experimentally recorded pair of data points from spike trains from 2 random variables M and N which are modelled as being taken from joint probability distribution $p_{(M,N)}(m, n)$. The density estimation approach revolves around the idea that for every point, probability mass function is estimated using a ball of defined volume around that point, which is given by the smoothing parameter h. Thus higher the measure of volume, the more precise is the estimation of count of points in ball, however the presumption that probability mass function $p_{(M,N)}(m, n)$ is approximately constant on ball, R becomes lesser precise.

For a point (m_i, n_i), the closest M spike train intervals to m_i and the nearest N spike train intervals to v_i is given by [6]

$$\begin{aligned} C_M(m_i, n_i) &= \{(m_j, n_j); d(m_j, m_i) \text{ is one of the smallest } M- \text{ distances })\} \\ C_N(m_i, n_i) &= \{(m_j, n_j) : d(n_j, n_i) \text{ is one of the smallest } N- \text{ distances })\} \end{aligned} \tag{3.34}$$

where the distance between spike trains, d is calculated using either of the following

methods:

- Euclidean Distance: The distance between points m and n in the Euclidean metric space is defined by

$$\begin{aligned} d(m,n) &= \sqrt{(n_1-m_1)^2+(n_2-m_2)^2+\cdots+(n_n-m_n)^2} \\ &= \sqrt{\sum_{i=1}^{n}(n_i-m_i)^2} \end{aligned} \tag{3.35}$$

- von-Rossum metric Primarily, the spike train data in the discrete form is converted into continuous functions using an exponential kernel [48]. The waveform that results from this is then used for the calculation of von-Rossum distance which is defined by

$$d(m,n) = \sum_{i,j} e^{-|m_i-m_j|/r} + \sum_{i,j} e^{-|n_i-n_j|/r} - 2\sum_{i,j} e^{-|m_i-n_i|/r} \tag{3.36}$$

  where the accuracy of spike times is determined by the time scale τ. Often, $\tau = 15ms$ is used for the calculation of distance. Though the proximity of the data points for the calculation of mutual information would use this metric, no matter what choice value of τ is chosen, the estimation mutual information remains insensitive to this parameter.

- Victor-Purpura distance: The distance between spike trains by Victor and Purpura determines the separation between 2 spike trains in a way such that minimal cost is incurred during the transformation of one spike train to other. This involves 3 basic operations spike insertion, deletion, and shift by an interval Δt [49].

Furthermore, the ball corresponding to the point (m_i, n_i) is given by:

$$C(m_i,n_i) = C_M(m_i,n_i) \cup C_N(m_i,n_i) \tag{3.37}$$

If, $C(m_i,n_i)$ be the count of points in $C(u_i,v_i)$, then $\#C(m_i,n_i)$ is given by [6]

$$\#C(m_i,n_i) = \#\left[C_M(m_i,n_i) \cap C_N(m_i,n_i)\right] \tag{3.38}$$

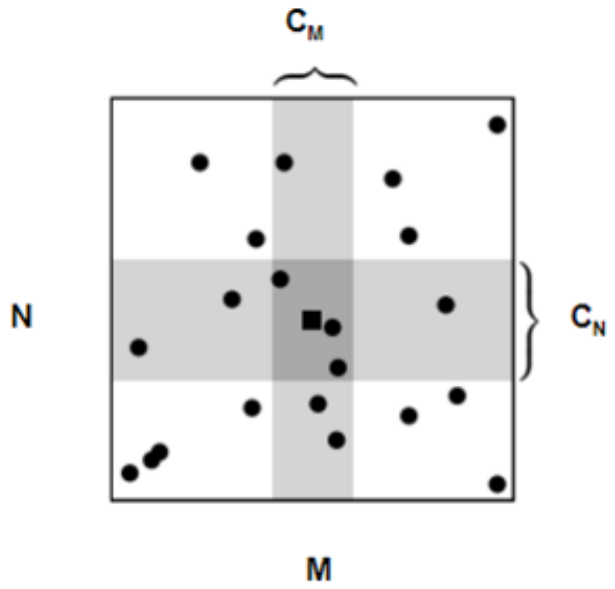

Figure 3.3: Cartoon to explain the working of KL-based method [6]

Figure **??** demonstrates how $\#\mathcal{C}\left(m_i, n_i\right)$ is calculated. This is a dummy figure and does not comply that the spaces have a single dimension, as the dimension of the spaces is undefined. The M space correspond to the closest M data points for a single pair (depicted by a square) of spike train interval. Similarly, the N space corresponds to the N closest data points, while, the other data points do not belong to either of the spaces. The darkly shaded region corresponds to the points that belong to both the spaces and the count of which contributes to the calculation of mutual information. This can be simply thought of as the K-nearest neighbors (KNN).

Thus, for each point (m_i, n_i), the set $\mathrm{C_M}\left(\mathrm{m_i}, \mathrm{n_i}\right)$ contains the point (m_i, n_i) itself and along with this it contains the other $h-1$ points that are closest when m_i is compared to m_j. Similarly, the set $C_N(mi, ni)$ contains closest $h-1$ points when n_i is compared to n_j and the $\#C(m_i, n_i)$ is the no of points that form the intersection. Thus, we can further easily estimate the mutual information with the Kozachenko-Leonenko approximation [6] given by :

$$I(M;N) \approx I_{KL}(P;h) = \frac{1}{n}\sum_{i=1}^{n} \log_2 n \frac{\#\left[C\left(m_i, n_i\right)\right]}{h^2} \tag{3.39}$$

When the distributions are independent, for varying values of r, the probability that $\#C\left(m_i, n_i\right) = r$ can be calculated. Selecting $h1$ points in $\mathrm{C_M\,(m_i n_i)}$ and not (m, n) itself randomly chooses $h1$ points out of $n1$ and calculating r determines the number of data points in $\mathrm{C_N\,(m_i n_i)}$; this gives

$$\mathrm{prob}\left(\#C\left(m, n_i\right) = r\right) = \frac{\left(\begin{array}{c} h-1 \\ r-1 \end{array}\right)\left(\begin{array}{c} n-h \\ h-r \end{array}\right)}{\left(\begin{array}{c} n-1 \\ h-1 \end{array}\right)} \tag{3.40}$$

Hence,

$$I_0(n, h) = \sum_{r=1}^{h} \mathrm{prob}\left(\#C\left(m_i, n_i\right) = r\right) \log_2 \frac{nr}{h^2} \tag{3.41}$$

Thus, an estimation of mutual information [6] is provided when the distributions are independent.

This produces an upward bias since the ball R will not necessarily contain absolutely $\#B$ points. Additionally, I_0 provides a straightforward formula for this bias depending only on the number of pairs, n and the smoothing parameter h. As h approached n, this bias gets removed but if not so, then this bias can be removed from the estimation of mutual information [6]:

$$I(M; N) \approx I(P; h) = I_{KL}(P; h) - I_0(n, h) \tag{3.42}$$

This provides a final estimation of the mutual information between two EMG channels.

Chapter 4

Data

4.1 Simulated Data

Fictitious data was created with the help of MATLAB to have simulation similar to that of the experimental data. Two arrays, A and B of size 7 X 8192 were created using the MATLABs randn() function, which generates normally distributed random numbers with 0 mean and 1 standard deviation for a given input size, n. Such array sizes were taken so that they correspond to the 7 trials and 8192 data points of the experimental data which was recorded for 4s (at 2000 Hz).

To verify that the estimation of mutual information using either of the methods is correct, the two arrays, A and B, needed to be correlated such that incrementing the correlation would increase the mutual information of the 2 entities. Hence, to do so, a third array was generated using the 2 arrays using the following [50],

$$C = \theta A + (1 - \theta)B \tag{4.1}$$

where θ corresponds to the correlation coefficient. The value of θ is varied between 0 and 1.

4.2 Experimental Data

Collection and maintenance of the EMG were performed by the Trinity Biomedical Science Institute and is solely provided for this research. The anonymity in identities

of the subject within the given data resolves the privacy concerns that relate to this study. Surface EMG from 8 muscles was recorded while the subjects were asked to perform a motor task to sit in front of a display screen. Sensors were employed to capture the force of the grip (on the chair arm) and the eye movements performed by the healthy subjects. This EMG was filtered from 0 Hz to 50 Hz and thereafter, was digitized at 2048 Hz.

EMG was taken from 8 subjects with the help of electrodes that capture the electrical signals in nerve conduction. This data is captured for 4 seconds with a sampling frequency of 2048Hz, thus generating a data of 8192 temporal values for each trial. Seven such trials are recorded for each patient.

The experimental data were recorded from the right hand of the on the 2 muscular regions First dorsal interosseous(FDI) and Abductor Pollicis Brevis(APB) (see Fig. 4.1). The FDI muscles are present at the back-side of the hand (between the thumb and the index finger), while the APB muscles reside on the palm under the thumbs. The EMG channel C5 corresponds to the neural signals recording for the FDI muscle, whereas the channel C6 corresponds to channel APB muscle. The overall dimensionality of the data used in the study is $2 \times 7 \times 8192$.

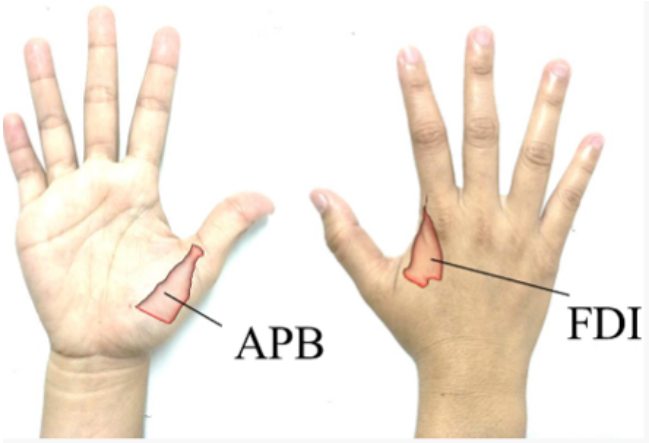

Figure 4.1: APB and FDI muscles of hand [7]

4.3 Decimation

Decimation is used to decrease the sample rate of the input sequence, thus, producing an approximation of the signal [51]. Both the fictitious and the experimental data are downsampled by the rate - 8.

Decimation is performed using the following MATLAB function:

$$\text{y = decimate(x,r)}$$

where x Is the input signal and r is the rate at which downsampling need to be performed [52]

Conducting decimation on each EMG channel for all the 7 trials involves the usage of IIR filter which reduces the sampling frequency to 256Hz. After downsampling the samples from the experimental and fictitious data, these are employed for the estimation of mutual information using the Kullback-Leibler and bin-based approach.

Chapter 5

Results

This section provides a discussion on the estimations of MI using the conventional bin-based approach and the innovative density based method with the help of simulated and experimental data. Furthermore, the effect of parameters: h and bins respective to the 2 approaches are studied. Lastly, the estimated mutual information is compared to information-theoretic linear measures - Power Spectral Density and Coherence.

5.1 Evaluation of MI using Simulated Data

Mutual Information is calculated for simulated data having a dimensionality of $2 \times 7 \times 8192$ data values. This corresponds to 4 seconds of data recorded per trial for two spike channels at a sampling frequency of 2048Hz. Seven such trials are considered here.

Primarily, we estimate MI using the density based approach, post which, we compare it to the bin-based approach and understand the effect of parameters

5.1.1 MI Calculation using Density(or KL)-based KNN approach

As already discussed, the Kullback Leibler-based approach evaluates mutual information by considering a region (typically a ball) around each data value. Hence, number of data values that fall within the region of each particular point contribute to the final estimate of mutual information . The size of this region depends on the quantity h (usually the radius) which determines the span of the point values. If the expanse of

this span is large, then more data points fall within the region, which can eventually lead to biased estimates, and similarly, smaller spans also generates inaccurate results. Hence, it is important to find an optimal value of h.

In order to estimate mutual information using fictitious data, 2 random vectors (generated from a normal distribution with mean 0 and variance 1) were considered. From this, a third vector was generated by varying the correlation from 0 to 1, with 0 depicting the least interdependecy while 1 showing highly dependence. Calculation of mutual information between the first and the third vectors was performed by varying values of correlation. Numerous such experiments were conducted to determine an optimal value of h. It was observed that the lower estimations of the MI are more closer to 0. As the value of h is increased, the estimations seem to rise to provide the optimal measurements (highlighted in blue in Fig 5.1) of the experiment.

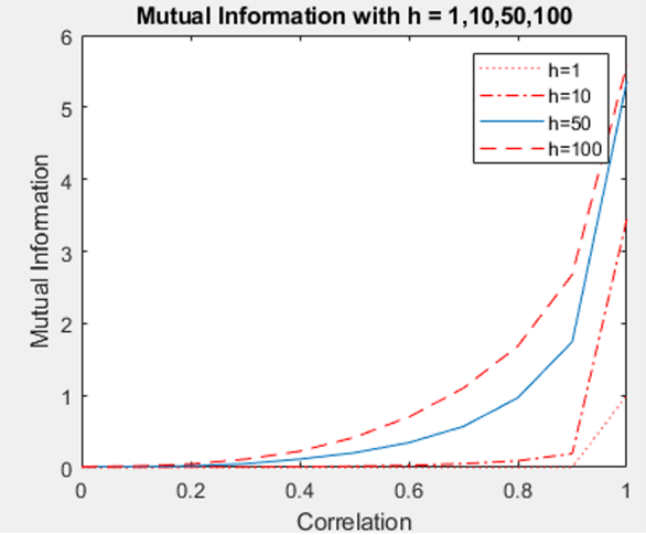

Figure 5.1: Mutual information using KL-based approach for varying values of h

5.1.2 Cross-validation with the bin-based method and the effect of h/bins

To validate the results from the KL approach, the conventional Bin-based approach was employed on the simulated vectors. As already discussed, this method takes into account the count of bins that correspond to a unit square in two dimensions. By

summing up data points lying in the bins, an estimate for the Akaike Information Criteria is obtained, which is used for the MI calculations. For conducting experiments, varying counts of bins were taken between the range (0,50).

With rising in the correlation between the two chosen vectors, results from the bin-based approach showed that when an increasing number of bins are used, there is a gradual increase in the MI estimates. If the bin count reaches a sufficiently large value, optimal estimates for the mutual information can be expected. However, further incrementing the count leads to deviation from the estimates and results in a condition commonly known as overfitting. On the contrary, lowering bin counts poorly estimate the mutual information (underfitting). Although precise estimations using this approach requires larger quantities of data, this study takes into account only the same amount of data as used in the KL(or density) method.

A comparison between the two approaches can be observed in 5.2. Lower values of h correspond to estimates of mutual information that are almost horizontal to the x-axis. On the other hand, on taking smaller number of bins, the Bin-based approach provides a fairly better estimate. Further incrementing the value of h and bins shows a rise in the mutual information which is higher in comparison to their respective initial results. Optimal values for MI are observed when the h-value is 50, which is seen in the bin-based approach when the bin-count is 40. It must be noted that the scales for the y-axis of the two plots vary. Increasing the values of parameter deviate the results from the optimal ones (in blue) in both the approaches. Moreover, the estimates from the density-based method show a quick rise in comparison to the ones from the bin-based method, which experience a gradual increase. This clearly suggests that the density-based method requires lesser amounts of data.

To firmly establish the findings of the KL-based approach (Table 5.1), mean cross-difference is calculated between the results of the KL-based approach and the optimal estimates from the bin-based method. This is given by the Eq. 5.1

$$\text{mean}\left(\frac{R_{KL} - O_B}{R_{KL} + O_B}\right) \tag{5.1}$$

where R_KL show the MI estimations from the KL approach and O_B denotes the optimal MI results from the bin-based method. Again here, it is evident that the lower values of the smoothing parameter show extreme values. For the simulated vectors,

Value of Smoothing parameter, h	Mean cross-difference between KL results and optimal Bin-based estimate (bins=40)
1	0.9595
10	0.8350
50	0.2692
100	0.3451

Table 5.1: Cross difference to compare Bin-based and KL-based based estimates from Simulated data

the value 50 shows the least difference between the two approaches and therefore is chosen as the optimal value of h for the density-based approach.

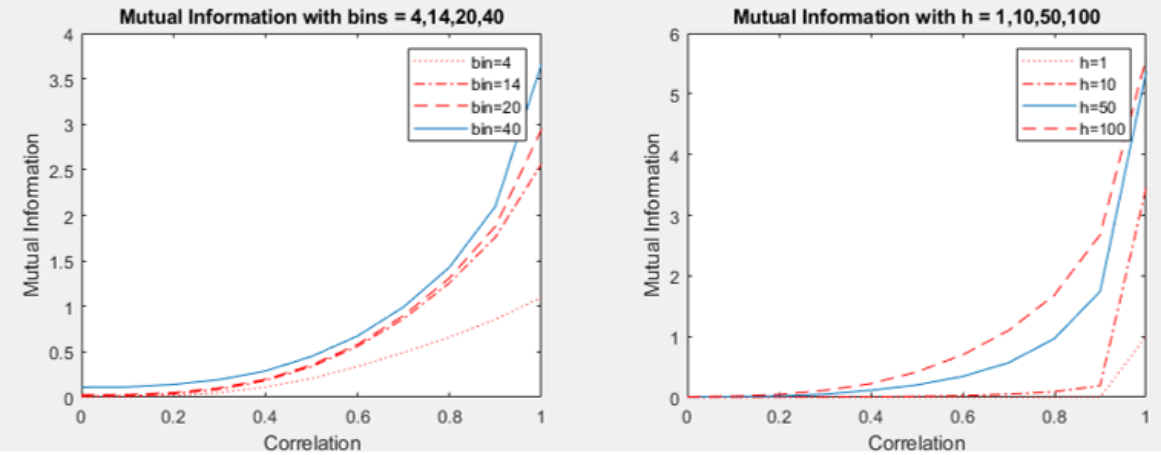

Figure 5.2: Comparing Bin-based method to the KL-based approach for different values of the parameters h/k

5.2 The effect of parameters

To understand the effect of parameters on electromyography, MI calculations were performed on the experimental data. Thus, this section provides discussions on the effect of smoothing parameter(h) on EMG, on optimal estimations using the KL based method, and on cross-validation of these results against the bin-based approach.

5.2.1 Experimenting with h and finding the optimum value

The experimental data consists of 2 channels that correspond to the APB and FDI muscles in our body. EMG Data for 8 healthy subjects is recorded for a period of 4 seconds with a sampling frequency of 2048Hz. Seven such trials are conducted on these subjects, thereby providing a dimensionality of 2X7X8192.

The density-based method is used to estimate the mutual information on this data. Primarily, auto-mutual information is calculated for the respective APB (C5) and FDI (C6) channels. Thereafter, cross-mutual information or simply the mutual information is estimated between the APB and FDI channels.

Auto-mutual information for the FDI and APB Channels

Auto-mutual information estimates the dependency of a point process (discrete-time event) with itself. It is a non-linear measure in correspondence to the linear information-theoretic measure − autocovariance and is used to understand the self-relationships in the muscular spike trains. Self-mutual information for the Channels C5 and C6 is conducted for the subject 103 to understand the effect of parameters. It must be noted that associations in neural spike trains are independent on the underlying structure of data. Hence, a different value of the smoothing parameter would be required to estimate the mutual information for each channel.

Calculation of mutual information in the channels is performed by lagging on the second signal. Lags of the order of 30 are considered for this study. Estimating mutual information for higher lags is restricted so that unnecessary complications in assessing a largely lagged signal could be curbed, as well as is done to lessen the computation time required to estimate the mutual information.

Fig 5.3 showcases the estimations of mutual information with Density-based method with the increasing values of h. MI results from these are validated against the bin-based approach (marked in blue in Fig 5.3). Smaller values of h lie closer to 0, hence it is difficult to observe them. However, If the value is increased a little, say to that of 100, then more sparse values of mutual information are observed. This is because the h-value is still lower than the optimal one. Upon further increasing the value to 1200, it is seen that MI estimates fall closer reside to those from the bin-based approach. Thus, this is observed as the best value of h for the C6−C6 MI estimate. Further

increasing h can lead to overfitting and this can be seen for h=4000.

A similar pattern is observed for the MI estimations of the C5 channel. Here MI is calculated for the h values 1, 100, 600, 1200 and 4000. As seen earlier, lower values generate estimates that are closer to 0 while significantly larger ones tend to overfit and fall farther away from the optimal estimates, thus providing poor results. Coincidentally for this channel, $h = 1200$ again generates MI results that are closer to those estimated from the optimal results from the bin-based and therefore, this value is chosen as the best value of the smoothing parameter for the auto-MI estimates for Channel C5.

Auto-mutual information results from both the C5 and the C6 channel take very high values for the initial lags. This is because of the experimental errors that arise while EMG is recorded. Hence, these initial estimations should be asserted more towards the recording flaws rather than counting them as the representing estimates of the actual neural synchrony.

(Cross-)mutual information between the 2 EMG channels- FDI and APB

The Cross-mutual information or simply mutual information is used to find the interactivity between the 2 EMG channels - C5 and C6 which are the respective spike trains measured from the FDI and APB muscles. In contrast to the estimates observed for auto-mutual information, the ones for the cross-MI show more variation along with the lags. However, the behavior of the cross-mutual information results closely correspond to those of the auto-mutual information and thus, smaller values of the parameter provide more sparse and deviating results while the larger one tends to over-fit and eventually diverge away from the optimal calculations. $h = 900$ show results that closer to the estimates from bin-based approach, hence, it is chosen as a befitting value of the parameter h to perform MI calculations C5 and C6.

5.2.2 Numerical Validation bin-based approach

To numerically validate the results, the mean-cross difference was calculated from the MI estimations using the experimental data. Table 5.2 shows the outcomes of this calculation with respect to each channel pair. For the C5-C6 pair, it was observed that $h = 900$ shows the lowest cross-difference against the optimal estimates from the

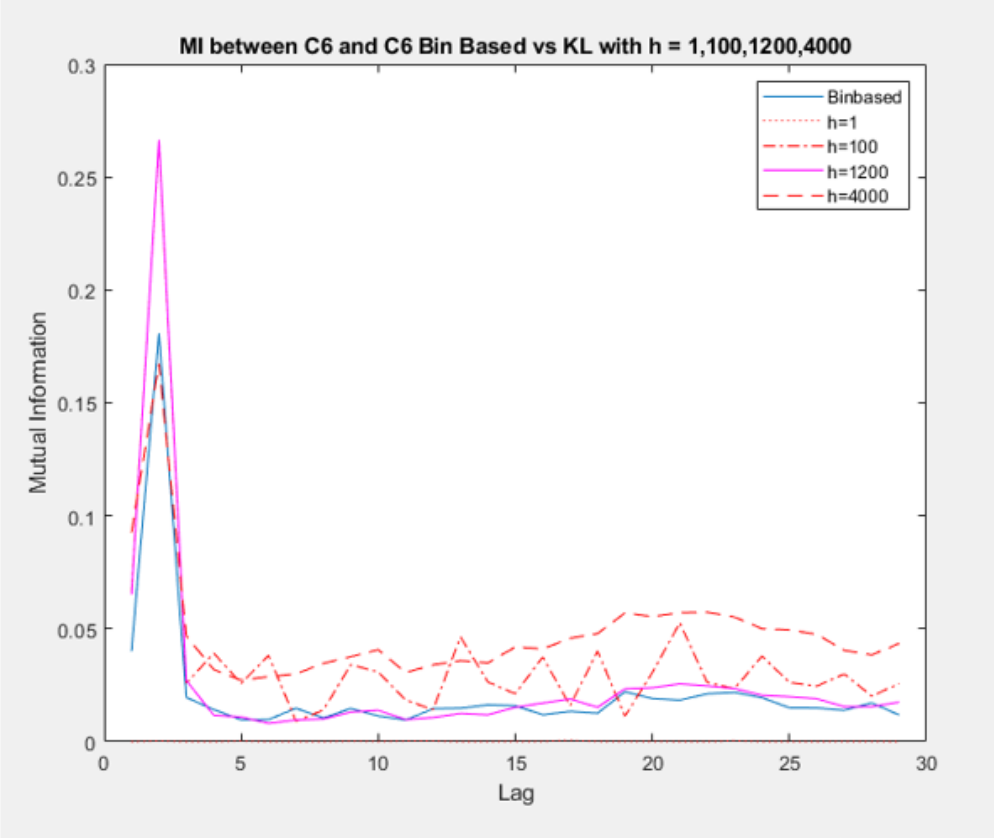

Figure 5.3: Auto-mutual information for FDI Channel with different values of h vs. KL-based estimates

bin-based approach. Hence, this value corresponds to the earlier chosen value of h, the conclusions for which were drawn from the plots. Similar outcomes can be drawn for the C5-C5 and C6-C6 pairs. Thus lower values underfit the cross-MI (or simply MI) estimations while the higher values overfit them.

5.2.3 Cross-comparison on the effect of parameters between the bin-based and KL-based apporaches

Numerous experiments were conducted by varying bin values for the 3 channel-pairs C5-C5, C6-C6, and C5-C6. Congruency could be observed between the results of these experiments and those of the simulated data which could help conclude the optimal count of bins for these channel-pairs. While the same bin count - 12 and h value - 1200

Channel-pair	**Value of smoothing parameter**	**Mean cross-difference between KL results and optimal Bin-based estimate (bins=40) for experimental data**
C5-C6	1	0.9845
C5-C6	100	0.3199
C5-C6	900	0.0941
C5-C6	1600	0.1194
C5-C6	2000	0.2064
C6-C6	1	0.9850
C6-C6	100	0.2940
C6-C6	1200	0.1071
C6-C6	4000	0.4543
C5-C5	1	0.9929
C5-C5	100	0.3356
C5-C5	800	0.2324
C5-C5	1200	0.1064
C5-C5	4000	0.2046

Table 5.2: Cross difference between Bin-based and KL-based based estimates using the Simulated data

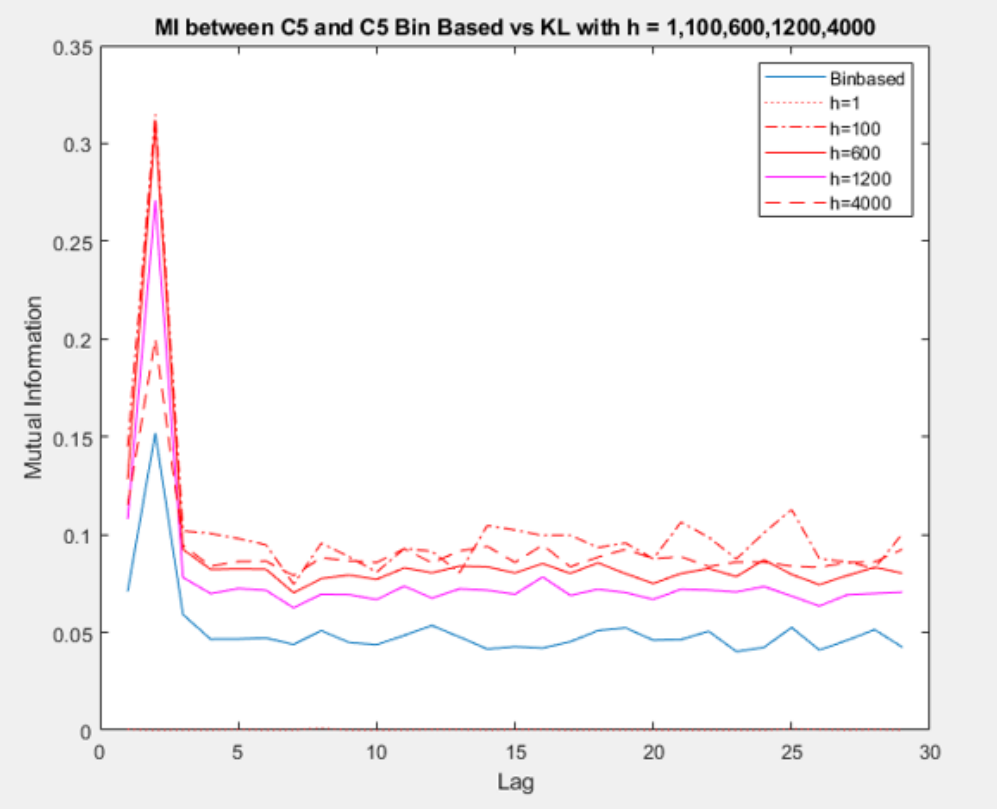

Figure 5.4: Auto-mutual information for APB Channel with different values of h vs. KL-based estimates

were recorded for auto-MI for channels C5 and C6, those for the MI between C5 and C6 took 14 and 900 respectively

5.3 Rythemic component in one EMG channel

Auto-mutual Information was calculated for the 2 EMG channels - C5 and C6 on 8 subjects. The evaluations were carried out by performing a lag of 256Hz. Rhythmic components in the MI estimates were discovered using the bin-based and the density-based approaches. The comparison of these patterns were carried out in the δ-band(0 to 4Hz),θ-band (4 to 8Hz), α-band(8 to 15Hz), β-band(15 to 30Hz) and γ-bands(greater than 30Hz) respectively . This section first explores the components observed from the estimates from both the approaches, which is followed by a discussion on the cross-

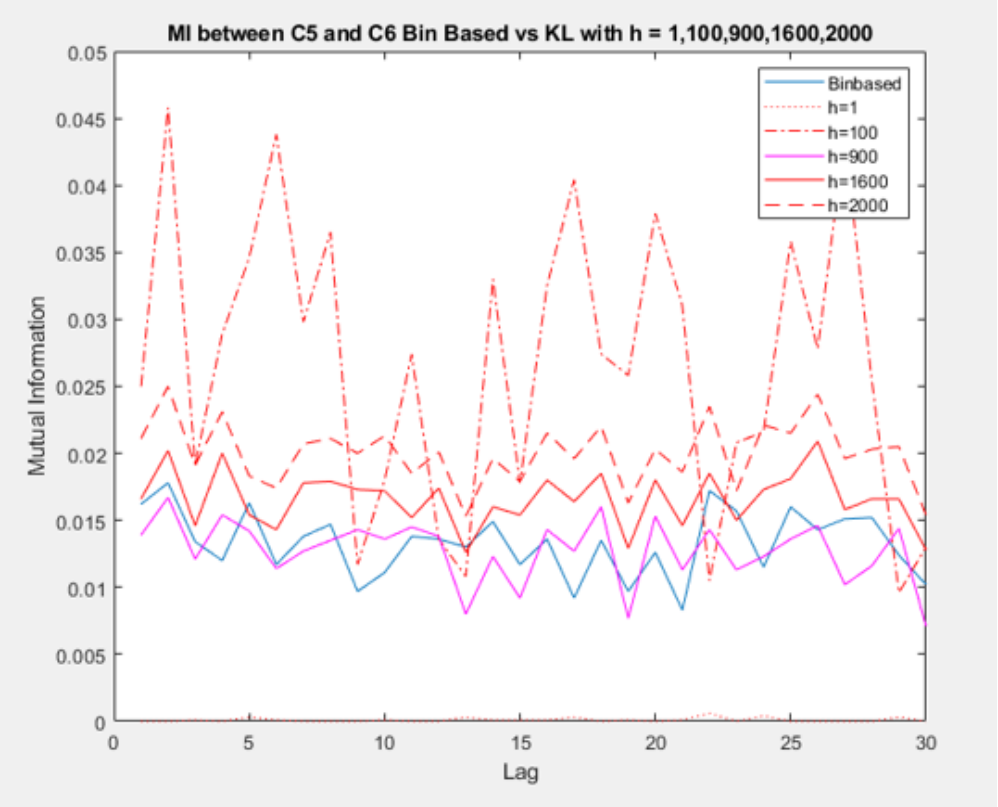

Figure 5.5: Mutual information between APB and FDI Channels with different values of h vs. KL-based estimates

comparison with the linear measure for auto-dependence: autospectrum.

5.3.1 Results from Lagged (Auto-)mutual information

Calculation of (auto-)mutual information using the 2 approaches was primarily conducted for the channel C6, post this, the methods were employed for the EMG channel corresponding to the APB muscle.

Decimation was performed on the data before the start of the experimentation. Data with the dimensionality of 7 X 8192 which corresponded to the sampling frequency of 2048Hz was down-sampled by a factor of 8, thereby, leading to the resultant frequency of 256Hz. Although data were recorded for a different no of trials for each subject, 7 trails were chosen with respect to each subject. Congruity in the number of trials was taken because the value of h goes inconsistent for the varying amounts of data. Optimal

bins for the bin-based approach were chosen as the result of numerous experiments that were conducted by performing variations in the bin count. Moreover, it must be noted that for a particular channel, the same value of h was used across all the subjects.

Smoothing by a factor of 5 was performed on the MI estimates using the MATLAB's y = smooth(x) function which uses a moving average filter on the data [52]. Larger estimates of mutual information were sliced out as these were more asserted to the experimental errors and hindered in the detailed analysis of the signal.

(Auto-)mutual information for Channel C6

Calculations of auto-mutual information are performed for the Channel C6 with itself. This is done keeping the first version of the channel intact, while a lag is performed on the second version. Lags of the order of 256Hz are conducted on the second version which corresponds to 1 second of data out of the 4 seconds that was recorded. The same value of the smoothing parameter, $h = 1200$ is used for all the subjects. A similarity in the rhythmic patterns is observed when the 2 approaches are compared to each other.

The beta-band is well known to show peaks in coherence plots when a force production task is conducted by the APB and FDI muscles. This shows interactivity between the nerve cells of muscles and motors neurons of the cortex (part of the brain responsible for movement) [53] As can be observed in almost all of the subjects (see Fig. 5.6,5.7,5.8), both the methods commonly show a larger spike in β-band. Such results are phenomenal to the estimation of mutual information using the density-based approach and signify the efficiency of this newly developed approach against the conventional one.

Close similarities can also be observed in the Gamma band, It is seen that results for some of the patients are lesser distorted in the bin-based approach than that from the density-based approach. A similar observation is made for the α band. However, it is difficult to closely examine these estimates given to the high distortions in the MI estimates in these frequency ranges.

Comparing results for the lower bands becomes difficult as the rhythmic patterns demonstrate much higher levels of contortions making it difficult to examine them by the naked eyes.

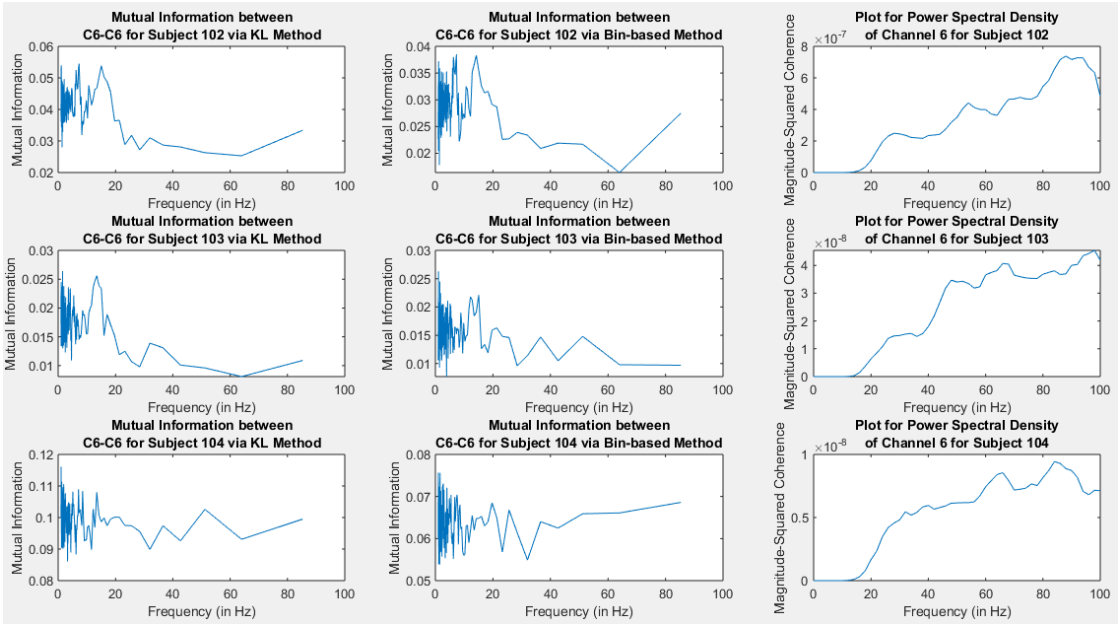

Figure 5.6: Cross-comparison of the Rythemic components in Channel 6 using 2 Mutual Information methods and Power Spectral for the subjects 102, 103, and 104

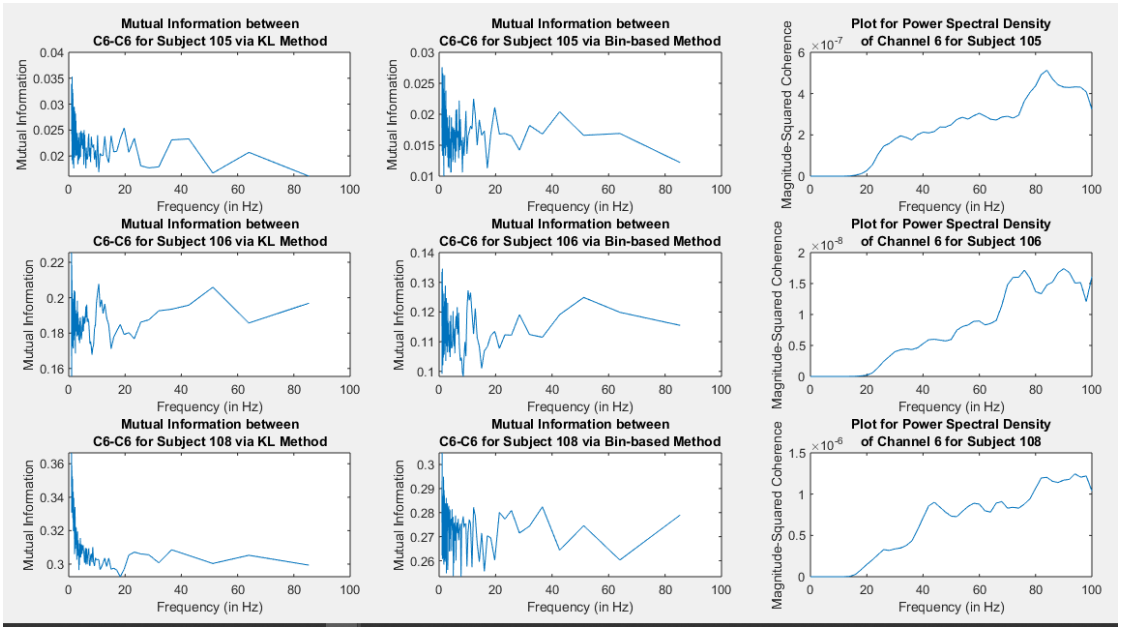

Figure 5.7: Cross-comparison of the Rythemic components in Channel 6 using 2 Mutual Information methods and Power Spectral for the subjects 105, 106, and 108

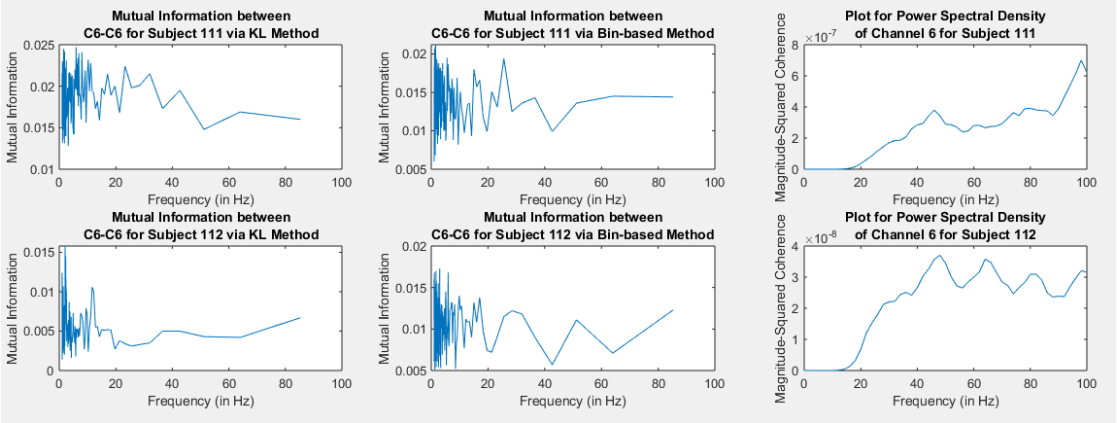

Figure 5.8: Cross-comparison of the Rythemic components in Channel 6 using 2 Mutual Information methods and Power Spectral for the subjects 111 and 112

(Auto-)mutual information for Channel C5

The calculation of auto-MI for the channel C5 is performed similarly as the channel C6. A comparison is made for the density-based approach against the bin-based method. A value of h=1200 is taken for the smoothing parameter against the bin count of 12. The behaviour of results (see Fig. 5.9, 5.10, 5.11) in C5 closely correspond to those are observed in C6. The β band that shows coherence between the activities of the cortex and the motor neurons again depicts peaks in its frequency range across all the subjects. Moreover, it is difficult to interpret the results in the smaller frequency ranges in comparison to γ band (having higher frequency band) which shows some levels of similarities between the estimates from the KL and the density-based method. But, a precise comparison in this is difficult as very few estimated of the mutual information fall within this frequency region.

5.3.2 Comparison to Power Spectral Density

Auto-MI estimations using the bin-based and density-based approach are compared against the linear measure: auto-spectrum also commonly known as the Power Spectral Density.

For both the channels C5(see Fig. 5.9, 5.10, 5.11) and C6 (see Fig. 5.6, 5.7, 5.8), it was observed that within the β-band the MI estimates from the two methods and the results of auto spectrum showed a similar trend. On the contrary, the results in the

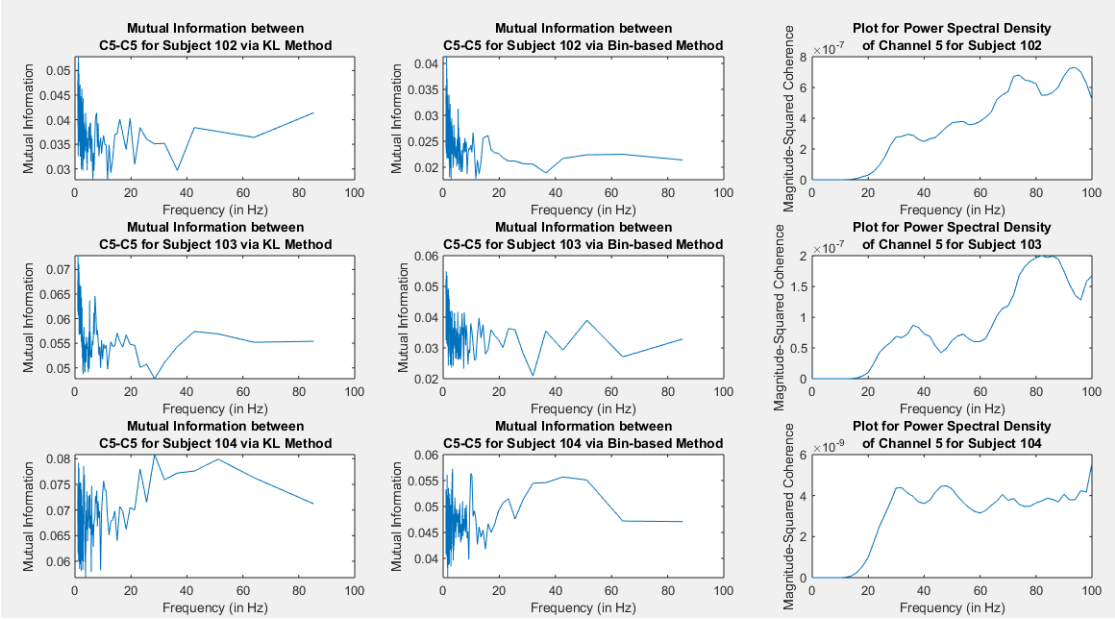

Figure 5.9: Cross-comparison of the Rythemic components in Channel 5 using 2 Mutual Information methods and Power Spectral for the subjects 102, 103, and 104

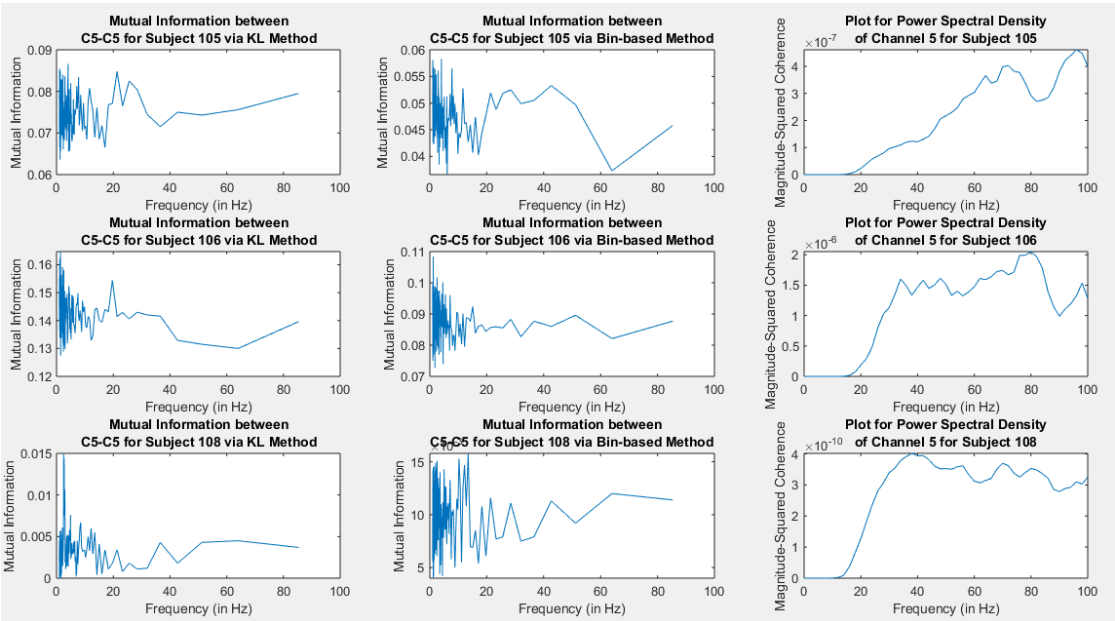

Figure 5.10: Cross-comparison of the Rythemic components in Channel 5 using 2 Mutual Information methods and Power Spectral for the subjects 105, 106, and 108

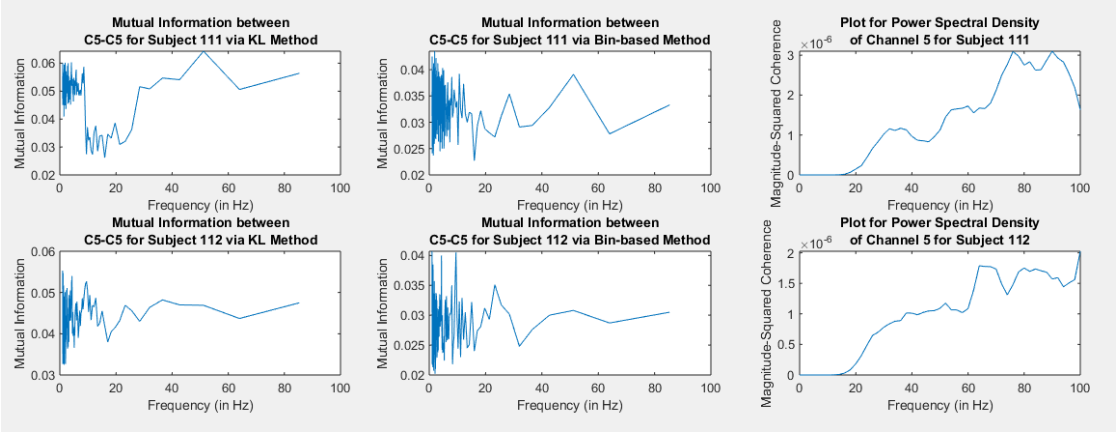

Figure 5.11: Cross-comparison of the Rythemic components in Channel 5 using 2 Mutual Information methods and Power Spectral for the subjects 111 and 112

gamma bands showed a dissimilar trend with some subjects. The δ and θ and α bands in the spectral density showed almost 0 mutual dependencies and hence it becomes difficult to compare them with the MI estimates. Thus, although the plots show vast dissimilarities between the two information theory measures, it must be noted that the auto spectrum is a poor estimator of the self-dependency in neural signals because the interactions within this type of data are non-linear.

5.4 Synchorny between 2 EMG Channels

Cross-Mutual information is calculated between the 2 EMG channels C5 and C6 to understand the synchrony between the APB and FDI muscles. This section provides an overview on how the calculation of the mutual information was performed between the 2 channels using the density-based and the bin-based methods, on the validation of these KL estimates to bin-based ones, and finally on the comparison of the MI estimates against the linear measure interdependency: Coherence.

5.4.1 Lagged (Cross-)Mutual Information for 2 EMG Channels

The cross mutual information was calculated between the channels C5 and C6 by lagging the samples of the latter channel by 256Hz. The optimal value of the smoothing

parameter was chosen from the earlier experiments conducted that used different variations of h (Ref. Chapter 5.2). The h value of 900 was taken for the Density-based method against the bin count of 14 for the Bin-based approach.

Comparing the estimates of MI with the Kl approach (see Fig. 5.12, 5.13, 5.14) against the binned one showcases interesting results. The β-band (15-30Hz) estimates for 2 methods show a similar trend across all of the patients, which shows peaks in the MI estimate in this interval. As previously discussed, it is known that spike activities in the β band show interactivity between the motor cortex neurons and nerve cells in the hand muscles. Such a pattern is easily captured by the conventional bin-based method. Interestingly, the results of MI using the density-based approach show similar trends across all the subjects. Similar trends in the γ (30Hz or higher) band can also be observed for the 2 approaches. While studying the δ (0-4Hz), θ (4-8Hz), and α (8-15Hz) bands is difficult, it can be fairly established that the results from the 2 approaches are closer to each other in these lower frequencies. Such similarities in the two approaches is an integral finding and suggest that the two approaches are coherent with each other.

5.4.2 Comparison to Spectral Coherence between 2 EMG Channels

The MI information estimates from the 2 approach are compared (see Fig. 5.12, 5.13, 5.14) to those of traditional linear approach: Coherence (or in this case the magnitude squared coherence). A similar trend in the beta band can be observed between the 3 estimates suggesting the MI estimates lie hand in hand to the coherence-based results. Such similar patterns are observed for the α, δ, and θ bands for most of the subjects with a few showing a little difference. However, such similarities in these frequency ranges cannot be strongly commented upon because of the disparities in the coherence results for some subjects. It is expected that shifting the signals $-$ 104 to the left and 106 to the right will establish similarities with the 2 subjects. Although at some level it can be said that results for the Gamma band also increase like those seen in the MI estimates, these outcomes cannot be firmly established given to the lesser amounts of estimates that fall in this region.

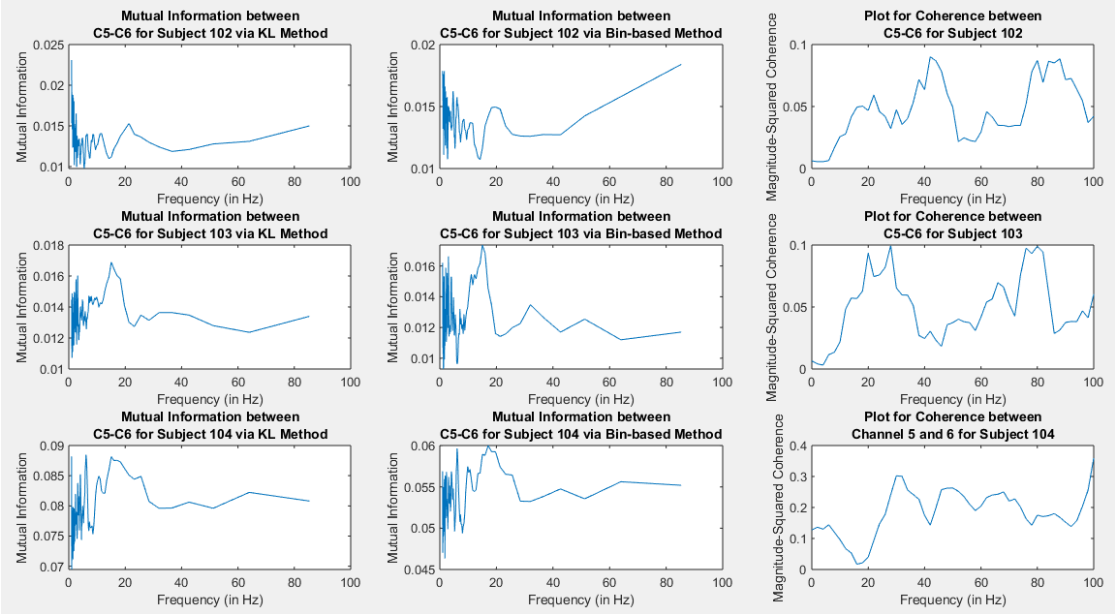

Figure 5.12: Cross-comparison of the Mutual Information calculated between Channel 5 and 6 against the coherence estimate for the subjects 102, 103, and 104

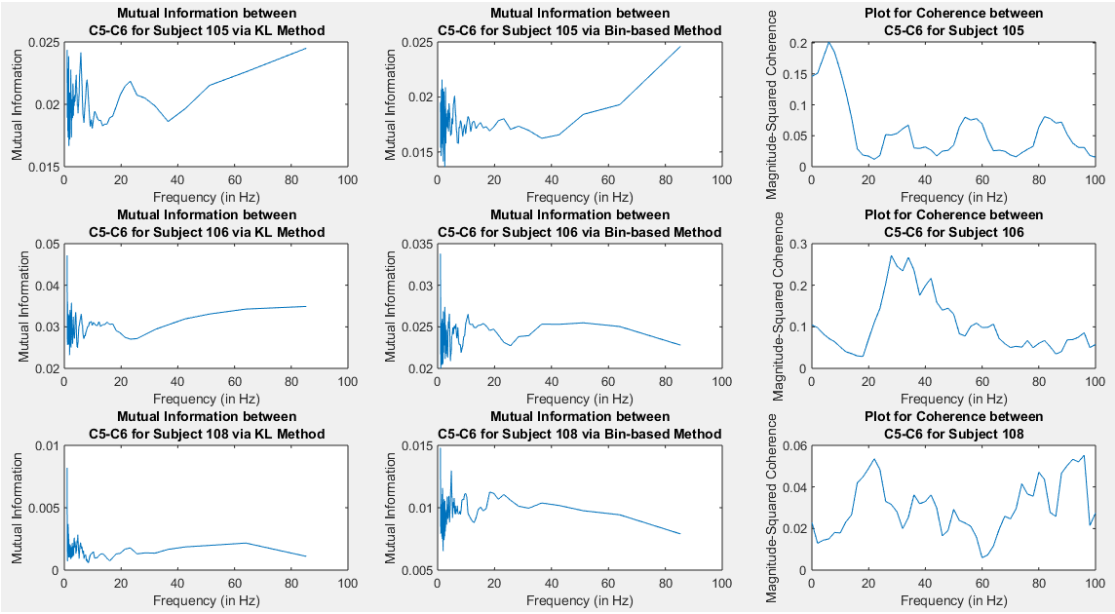

Figure 5.13: Cross-comparison of the Mutual Information calculated between Channel 5 and 6 against the coherence estimate for the subjects 105, 106, and 108

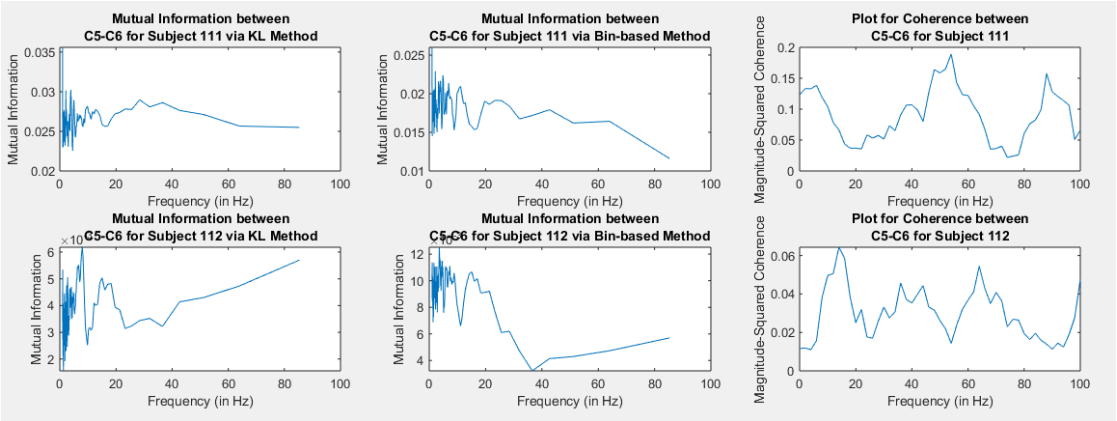

Figure 5.14: Cross-comparison of the Mutual Information calculated between Channel 5 and 6 against the coherence estimate for the subjects 111 and 112

Chapter 6

Discussions and conclusions

This section provides a discussion on the conclusions drawn from these results, describes the limitations and highlights the future work concerning the study.

6.1 Summary of New Contributions

The study involves the calculation of mutual information using the density-based approach and the bin-based method. Significant conclusions were drawn out from this work.

Primarily, on cross-comparison of the MI estimations from the bin-based approach to the density-based method, it was observed that results of the latter showed similar variation and took values that were in close correspondence to the bin-based approach. This validates the results from the newly established KL approach using the EMG data and establishes that the density-based approach is also efficient to measure the synchrony between neuromuscular signals.

On comparison of the results from the KL approach to that from the information-theoretic linear measure - Coherence and Autospectrum (or the PSD), it was observed that estimates from both the density-based and bin-based methods produced results that are similar to these measures. Although a close similarity was not observed, it must be noted that the coherence and PSD are traditional measures and poorly estimate the inter-dependencies between the two muscular regions.

6.2 Applications in neuroscience and neurological diagnostics

Intermuscular and neuromuscular dependencies measured with the traditional linear approach are more sparse and generate inaccurate results as these inherently depend upon the linearity of the data, which is not the case for the neural signals. Thus, these measures are less reliable

Moreover, for the estimations of the non-linear measure - the mutual information, the conventional methods require tremendously large quantities of data. Such huge quantities are infeasible to measure but are needed for the precise estimations from these approaches.

The density-based approach (used in this study) requires significantly lesser amounts of data to achieve precise evaluations. This can be significant in the neuroscience domain and can be used for advanced studies of inter-muscular and neuromuscular regions and thus can better aid in the diagnosis of nerve-muscle and muscle-muscle related disorders like the ALS and for a genetic muscular disorder like the Duchene's muscular dystrophy. Apart from this, potential use lies in designing advanced prosthesis.

6.3 Limitations

Though the innovative KL(or density) based approach is advantageous, it comes with several limitations.

Primarily, estimating MI using this approach is computationally expensive. This comes from iteratively finding values within closer proximity for each paired-data value of the input signal. This disadvantage would limit its usage in the studies that require a computationally efficient environment.

Moreover, it becomes difficult to distinguish between the mutual information estimates if the values of h are closer to each other. The MI values coincide with each other when such smaller values of h are taken. This hinders in the process of choosing an optimal value of the smoothing parameter, h.

Apart from this, for the varying amounts of data, the value of smoothing parameter varies, hence this again makes it challenging to determine the best measure of h. Thus

it can be established that choosing the value of the smoothing parameter is a major challenge in this approach.

6.4 Future Work

Although appropriate values of h were taken in this study by performing cross-validation against the bin-based approach, a strong measure needs to be devised to determine the optimal value of h. Moreover, advancements that increase the computational efficiency of the algorithm can also be invented.

Lastly, as used in this dissertation, to study the relationships between two muscular regions, this approach can also be employed for understanding electrical conductivity in the intercortical regions which can be used explain neural synchrony within the brain.

Bibliography

[1] E. N. Brown, R. E. Kass, and P. P. Mitra, "Multiple neural spike train data analysis: state-of-the-art and future challenges," *Nature neuroscience*, vol. 7, no. 5, p. 456, 2004.

[2] E. J. Kobylarz, A. E. Hudson, A. Kamal, R. J. DeBellis, and N. D. Schiff, "Power spectrum and coherence analysis of the electroencephalogram from two minimally conscious patients with severe asymmetric brain damage," in *Society for Neuroscience 35th Annual Meeting, Abstract*, 2005.

[3] C. J. De Luca, A. Adam, R. Wotiz, L. D. Gilmore, and S. H. Nawab, "Decomposition of surface emg signals," *Journal of neurophysiology*, vol. 96, no. 3, pp. 1646–1657, 2006.

[4] D. C. Preston and B. E. Shapiro, *Electromyography and Neuromuscular Disorders E-Book: Clinical-Electrophysiologic Correlations (Expert Consult-Online and Print)*. Elsevier Health Sciences, 2012.

[5] N. M. Timme and C. Lapish, "A tutorial for information theory in neuroscience," *eNeuro*, vol. 5, no. 3, 2018.

[6] C. Houghton, "Calculating the mutual information between two spike trains," *Neural computation*, vol. 31, no. 2, pp. 330–343, 2019.

[7] S. Jitaree and P. Phukpattaranont, "Force classification using surface electromyography from various object lengths and wrist postures," *Signal, Image and Video Processing*, pp. 1–8, 2019.

[8] A. Borst and F. E. Theunissen, "Information theory and neural coding," *Nature neuroscience*, vol. 2, no. 11, p. 947, 1999.

[9] A.-H. Shapira and I. Nelken, "Binless estimation of mutual information in metric spaces," *Spike Timing: Mechanisms and Function; DiLorenzo, P., Victor, J., Eds*, pp. 121–135, 2013.

[10] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical review E*, vol. 69, no. 6, p. 066138, 2004.

[11] C. D. Brody, "Correlations without synchrony," *Neural computation*, vol. 11, no. 7, pp. 1537–1551, 1999.

[12] D. R. Brillinger, "Nerve cell spike train data analysis: a progression of technique," *Journal of the American Statistical Association*, vol. 87, no. 418, pp. 260–271, 1992.

[13] R. E. Kass, V. Ventura, and C. Cai, "Statistical smoothing of neuronal data," *Network-Computation in Neural Systems*, vol. 14, no. 1, pp. 5–16, 2003.

[14] A. Aertsen, G. Gerstein, M. Habib, and G. Palm, "Dynamics of neuronal firing correlation: modulation of" effective connectivity"," *Journal of neurophysiology*, vol. 61, no. 5, pp. 900–917, 1989.

[15] H. Ito and S. Tsuji, "Model dependence in quantification of spike interdependence by joint peri-stimulus time histogram," *Neural computation*, vol. 12, no. 1, pp. 195–217, 2000.

[16] M. N. Shadlen and W. T. Newsome, "The variable discharge of cortical neurons: implications for connectivity, computation, and information coding," *Journal of neuroscience*, vol. 18, no. 10, pp. 3870–3896, 1998.

[17] R. Gütig, A. Aertsen, and S. Rotter, "Statistical significance of coincident spikes: count-based versus rate-based statistics," *Neural Computation*, vol. 14, no. 1, pp. 121–153, 2002.

[18] D. B. Percival and A. T. Walden, *Wavelet methods for time series analysis*, vol. 4. Cambridge university press, 2000.

[19] B. Pesaran, J. S. Pezaris, M. Sahani, P. P. Mitra, and R. A. Andersen, "Temporal structure in neuronal activity during working memory in macaque parietal cortex," *Nature neuroscience*, vol. 5, no. 8, p. 805, 2002.

[20] T. Kreuz, J. S. Haas, A. Morelli, H. D. Abarbanel, and A. Politi, "Measuring spike train synchrony," *Journal of neuroscience methods*, vol. 165, no. 1, pp. 151–161, 2007.

[21] T. Kreuz, D. Chicharro, C. Houghton, R. G. Andrzejak, and F. Mormann, "Monitoring spike train synchrony," *Journal of neurophysiology*, vol. 109, no. 5, pp. 1457–1472, 2012.

[22] C. M. Gómez, E. I. Rodríguez-Martínez, A. Fernández, F. Maestú, J. Poza, and C. Gómez, "Absolute power spectral density changes in the magnetoencephalographic activity during the transition from childhood to adulthood," *Brain topography*, vol. 30, no. 1, pp. 87–97, 2017.

[23] D. S. Reich, F. Mechler, and J. D. Victor, "Independent and redundant information in nearby cortical neurons," *Science*, vol. 294, no. 5551, pp. 2566–2568, 2001.

[24] C. J. Rozell and D. H. Johnson, "Examining methods for estimating mutual information in spiking neural systems," *Neurocomputing*, vol. 65, pp. 429–434, 2005.

[25] Y. Masamizu, Y. R. Tanaka, Y. H. Tanaka, R. Hira, F. Ohkubo, K. Kitamura, Y. Isomura, T. Okada, and M. Matsuzaki, "Two distinct layer-specific dynamics of cortical ensembles during learning of a motor task," *Nature neuroscience*, vol. 17, no. 7, p. 987, 2014.

[26] C. Houghton, "Calculating mutual information for spike trains and other data with distances but no coordinates," *Royal Society open science*, vol. 2, no. 5, p. 140391, 2015.

[27] G. E. Robertson, G. E. Caldwell, J. Hamill, G. Kamen, and S. Whittlesey, *Research methods in biomechanics*. Human kinetics, 2013.

[28] Y. Wu, M. Á. M. Martínez, and P. O. Balaguer, "Overview of the application of emg recording in the diagnosis and approach of neurological disorders," in *Electrodiagnosis in New Frontiers of Clinical Research*, IntechOpen, 2013.

[29] R. Merletti and D. Farina, "Analysis of intramuscular electromyogram signals," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1887, pp. 357–368, 2008.

[30] S. Podnar and M. Mrkaić, "Size of motor unit potential sample," *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, vol. 27, no. 2, pp. 196–201, 2003.

[31] L. Reinstein, F. Twardzik, and J. K. Mech, "Pneumothorax: a complication of needle electromyography of the supraspinatus muscle.," *Archives of physical medicine and rehabilitation*, vol. 68, no. 9, pp. 561–562, 1987.

[32] S. L. Pullman, D. S. Goodin, A. I. Marquinez, S. Tabbal, and M. Rubin, "Clinical utility of surface emg: report of the therapeutics and technology assessment subcommittee of the american academy of neurology," *Neurology*, vol. 55, no. 2, pp. 171–177, 2000.

[33] M. J. Zwarts, G. Drost, and D. F. Stegeman, "Recent progress in the diagnostic use of surface emg for neurological diseases," *Journal of electromyography and kinesiology*, vol. 10, no. 5, pp. 287–291, 2000.

[34] G. D. Meekins, Y. So, and D. Quan, "American association of neuromuscular & electrodiagnostic medicine evidenced-based review: Use of surface electromyography in the diagnosis and study of neuromuscular disorders," *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, vol. 38, no. 4, pp. 1219–1224, 2008.

[35] S. Kattla and M. M. Lowery, "Fatigue related changes in electromyographic coherence between synergistic hand muscles," *Experimental brain research*, vol. 202, no. 1, pp. 89–99, 2010.

[36] K. I. Shaikhly, F. B. Hamdan, and F. S. Al-Ani, "Power spectrum analysis and conventional electromyogram in duchenne muscular dystrophy," *Saudi medical journal*, vol. 21, no. 11, pp. 1038–1042, 2000.

[37] X. Li, H. Shin, P. Zhou, X. Niu, J. Liu, and W. Z. Rymer, "Power spectral analysis of surface electromyography (emg) at matched contraction levels of the first dorsal interosseous muscle in stroke survivors," *Clinical Neurophysiology*, vol. 125, no. 5, pp. 988–994, 2014.

[38] J. F. Alonso, M. A. Mañanas, D. Hoyer, Z. L. Topor, and E. N. Bruce, "Evaluation of respiratory muscles activity by means of cross mutual information function at different levels of ventilatory effort," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 9, pp. 1573–1582, 2007.

[39] R. N. Khushaba and S. Kodagoda, "Electromyogram (emg) feature reduction using mutual components analysis for multifunction prosthetic fingers control," in *2012 12th International Conference on Control Automation Robotics & Vision (ICARCV)*, pp. 1534–1539, IEEE, 2012.

[40] M. A. Guevara and M. Corsi-Cabrera, "Eeg coherence or eeg correlation?," *International Journal of Psychophysiology*, vol. 23, no. 3, pp. 145–153, 1996.

[41] J. J. Shynk, *Probability, random variables, and random processes: theory and signal processing applications*. John Wiley & Sons, 2012.

[42] B. Bhushan, "Tribology," *Mcchicr of Muptic Stomgc Devices*, 1990.

[43] D. M. Halliday and J. R. Rosenberg, "Time and frequency domain analysis of spike train and time series data," in *Modern techniques in neuroscience research*, pp. 503–543, Springer, 1999.

[44] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[45] S. P. Strong, R. Koberle, R. R. D. R. Van Steveninck, and W. Bialek, "Entropy and information in neural spike trains," *Physical review letters*, vol. 80, no. 1, p. 197, 1998.

[46] R. B. Nelsen, *An introduction to copulas*. Springer Science & Business Media, 2007.

[47] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected papers of hirotugu akaike*, pp. 199–213, Springer, 1998.

[48] M. v. Rossum, "A novel spike distance," *Neural computation*, vol. 13, no. 4, pp. 751–763, 2001.

[49] J. D. Victor and K. P. Purpura, "Metric-space analysis of spike trains: theory, algorithms and application," *Network: computation in neural systems*, vol. 8, no. 2, pp. 127–164, 1997.

[50] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[51] G. A. Baxes, *Digital image processing: principles and applications*. Wiley New York, 1994.

[52] MATLAB, *version 9.3.1 (R2018a)*. Natick, Massachusetts: The MathWorks Inc., 2018.

[53] A. Reyes, C. M. Laine, J. J. Kutch, and F. J. Valero-Cuevas, "Beta band corticomuscular drive reflects muscle coordination strategies," *Frontiers in computational neuroscience*, vol. 11, p. 17, 2017.

Appendix

Abbreviations

MI	Mutual Information
EMG	Electromyography
ALS	Amyotrophic Lateral Screlosis
FDI	First Dorsal Interosseous muscle
APB	Abductor pollicis brevis muscle
KL	Kozachenko-Leonenko Estimator
PSD	Power Spectral Density
OSAS	Obstructive sleep apnea syndrome