A COMPARATIVE ANALYSIS OF COHERENCE MEASURES FOR ELECTROENCEPHALOGRAPHY

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Electronics and Communication Engineering

by
Serhat ÇAĞDAŞ

December 2018
İZMİR

We approve the thesis of **Serhat ÇAĞDAŞ**

Examining Committee Members:

Prof. Dr. Bilge KARAÇALI
Department of Electrical and Electronics Engineering
İzmir Institute of Technology

Assoc. Prof. Şevket GÜMÜŞTEKİN
Department of Electrical and Electronics Engineering
İzmir Institute of Technology

Assoc. Prof. Devrim ÜNAY
Department of Biomedical Engineering
İzmir University of Economics

21 December 2018

Prof. Dr. Bilge KARAÇALI
Supervisor, Department of Electrical and Electronics Engineering
İzmir Institute of Technology

Prof. Dr. Enver TATLICIOĞLU
Head of the Department of
Electrical and Electronics Engineering

Prof. Dr. Aysun SOFUOĞLU
Dean of the Graduate School of
Engineering and Sciences

ACKNOWLEDGMENTS

This work was supported in part by the ÖYP programme of the Council of Higher Education and in part by the TÜBİTAK grant 117E784 awarded to Prof. Dr. Bilge KARAÇALI.

I am grateful to my supervisor Prof. Dr. Bilge KARAÇALI for not only his assistance along the writing process of the thesis but for his helpful advices about life and being a good academic.

I would like to thank Bilal Orkan OLCAY. Whenever I had a question he was there with a satisfying answer.

Additionally, I would like to thank Fatih GÜLEÇ for the sincere friendship and motivational coffee breaks.

Finally and most importantly, I would like to thank my family for the endless support they gave in any case.

ABSTRACT

A COMPARATIVE ANALYSIS OF COHERENCE MEASURES FOR ELECTROENCEPHALOGRAPHY

Functional connectivity is often used in brain-computer interface studies as well as other neuroscience fields as a feature extraction method. In the functional connectivity using electroencephalography (EEG), connectivity patterns are extracted by a dependency matrix showing the coherence between electrode pairs. A variety of dependence measures can be used to calculate this matrix. In this study, a total of 15 coherence measures were analyzed comparatively in terms of computation time, accuracy and statistical significance in discriminating motor/motor imagery activities. As dependence measures, in addition to methods used in the literature for brain connectivity, five other methods used as contrast function in independent component analysis and two novel mutual information calculators proposed in this study were evaluated. Furthermore, a novel hierarchical clustering based statistical test procedure was also proposed for motor/motor imagery activity comparison, along with a similar statistical significance test applied on data from 103 subjects on four different activity types. In experiments on real data set, significance results of dependence measures differed according to the type of activity and time window duration of activity signals. Considering both computation time and accuracy performances on synthetic data, a number of methods with high statistical significance and different dependence characteristics were identified as feasible for a connectivity based brain-computer interface.

ÖZET

ELEKTROENSEFALOGRAFİ İÇİN UYUMLULUK ÖLÇÜTLERİNİN KARŞILAŞTIRMALI ANALİZİ

Fonksiyonel bağıntılılık, diğer sinirbilim alanlarının yanı sıra beyin-bilgisayar arayüzü çalışmalarında da bir özellik çıkarım yöntemi olarak sıklıkla kullanılmaktadır. Elektroensefalografiyi (EEG) kullanan fonksiyonel bağıntılılık çalışmalarında, bağıntı desenleri elektrod çiftleri arasındaki uyumluluğu gösteren bir bağımlılık matrisi ile çıkarılmaktadır. Bu matrisin hesaplanmasında çeşitli uyumluluk ölçütleri kullanılmaktadır. Bu çalışmada, toplam 15 uyumluluk ölçütü hesaplama zamanı, doğruluk ve motor/hayali motor aktivitelerini ayırmadaki anlamlılığı yönünden karşılaştırmalı olarak analiz edilmiştir. Uyumluluk ölçütü olarak beyin bağıntılılığı literatüründe kullanılan yöntemlerin dışında, bağımsız bileşen analizi yönteminde kontrast fonksiyonu olarak kullanılan beş ölçüt ve bu çalışmada önerilen iki yeni ortak bilgi miktarı hesaplama yöntemi değerlendirilmiştir. Ayrıca, dört farklı aktivite türünde 103 denekten elde edilen veriye uygulanan bir istatistiksel anlamlılık testinin yanında, motor/hayali motor aktivite karşılaştırmaları için de, hiyerarşik kümelemeye dayalı yeni bir istatistiksel test prosedürü önerilmiştir. Gerçek veri setlerine uygulanan deneylerde kullanılan ölçütler, uygulandığı aktivite çeşidi ve zaman aralığına göre farklı anlamlılık değerleri vermiştir. Sentetik veri üzerinde gösterdikleri zaman ve doğruluk performansları da göz önünde bulundurularak istatistiksel anlamlılığı yüksek ve farklı bağıntılılık karakteristiği sunan birkaç yöntem, bağlantılılık temelli beyin-bilgisayar arayüzünde uygulanabilir olarak tanımlanmıştır.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

LIST OF ABBREVIATIONS

CNS . . . . . Central Nervous System
EEG . . . . . Electroencephalography
MEG . . . . . Magnetoencephalography
fMRI . . . . . Functional Magnetic Resonance Imaging
PET . . . . . Positron Emission Tomography
SPECT . . . . . Single Photon Emission Tomography
BCI . . . . . Brain-Computer Interface
ERP . . . . . Event Related Potential
SSVEP . . . . . Steady-State Visual Evoked Potential
ERS/ERD . . . . . Event Related Synchronization/ Desynchronization
ICA . . . . . Independent Component Analysis
BOLD . . . . . Blood-Oxygenation Level Dependent
CT . . . . . Computed Tomography
PLV . . . . . Phase Locking Value
PLI . . . . . Phase Lag Index
iCOH . . . . . imaginary Part of Coherence
SCC . . . . . Spectral Correlation Coefficient
SW . . . . . Schweizer-Wolff
MI . . . . . Mutual Information
UV . . . . . Unit Vector
CCA . . . . . Canonical Correlation Coefficient
KGV . . . . . Kernel Generalized Variance
KCCA . . . . . Kernel Canonical Correlation Coefficient
REGO . . . . . Rank-Based Graph Optimization
HiCl . . . . . Hierarchical Clustering

CHAPTER 1

INTRODUCTION

Brain, as the main part of central neural system (CNS), consists of a large number of neurons and manages all perceptual, motor and cognitive tasks by the complex network of these cells. Although there exist invasive applications, noninvasive brain imaging techniques are mostly used since its sensitive nature in a rigid skull. The techniques such as electroencephalography (EEG) and magnetoencephalography (MEG) receive electromagnetic signals caused by brain activity. Other techniques like functional magnetic resonance imaging (fMRI), positron emission tomography (PET) and single photon emission computerized tomography (SPECT) measure the brain state by benefiting from indirect features of it. Usually, spatial and temporal resolutions of these techniques are considered to decide on which are to use in studies. Especially EEG that measures the electrical signals over the scalp is utilized for functionality studies of the brain and brain computer interface (BCI) applications.

Brain computer interface studies aim to control the devices around users by the inputs of brain activity without muscular motion. As a system mostly constructed on EEG signals, BCI is predominantly based on selectivity of attention (also named as oddball paradigm) and neural feedback paradigms. Signals such as event related potential (ERP), steady-state visual evoked potential (SSVEP) and event related synchronization/desynchronization (ERS/ERD) are used for measuring external stimuli in these paradigms. Unlike the selectivity of attention, motor imagery based BCI studies do not need an external stimulus and the subject only imagines the movement to be performed. These studies usually consider ERS/ERD signals (Pfurtscheller and Neuper, 2001). But the success rate of motor imagery based BCI is relatively low despite its potential of information transfer speed. Moreover, it requires long neural feedback based training sessions (Graimann et al., 2010).

As a remedy to these problems, areas of the brain that activate together during the imaginary motor functions can be identified (Delorme et al., 2002), (Hamedi et al., 2016). Many signal processing methods are proposed to measure which channels or brain areas work together by analyzing EEG time series. These studies are categorized in the brain connectivity concept (Sakkalis, 2011). Among these studies, the primary methods proposed to measure functional connectivity are linear correlation (Adey et al., 1961),

wavelet coherence (Lachaux et al., 2002) and phase locking value (Rosenblum et al., 1996) methods. It should be emphasized that these methods can identify the areas working together to a certain extent in spite of artificial dependence caused by the conductivity properties of the scalp and the brain itself (volume conduction effect) (Srinivasan et al., 2007). In addition to these methods, several methods are proposed to eliminate the volume conductivity problem (Stam et al., 2007), (Nolte et al., 2004). Finally, mutual information based methods are also used as dependence measures (Sakkalis, 2011). But the lack of information about marginal and joint probability density functions create an obstacle for these methods.

Another problem for the dependence measure methods are time lag of communication between channels or brain areas. Furthermore, EEG signals are stationary only for short time intervals (declared as quasi-stationary) (Sanei and Chambers, 2007). This also causes an inconsistency for the methods that require stationarity.

Considering the above mentioned problems, different methods have differing sets of advantages and disadvantages on detecting connectivity patterns. The main purpose of this thesis is to carry out a comparative analysis of dependence measures used for functional connectivity or have the potential to be used to measure dependence for the evaluation of motor/motor imagery signals in BCI systems.

This thesis is organized as follows. The second section consists of a brief background about brain imaging techniques, brain computer interface and the brain connectivity concept. This section will help reader to understand why EEG is chosen for this study and provide a preliminary information for the rest of the study. The following section starts with an introduction to Physionet EEG motor/motor imagery dataset used in the experiments and the preprocessing procedures carried out on the dataset. After that, various dependence measure methods are described in detail. In this section, in addition to dependence measures already used for connectivity analysis, five other methods used in independent component analysis (ICA) technique that searches the linear transformation matrix to find independent sources are explored as potential dependence measures. Furthermore, two new information theoretic methods are proposed as mutual information estimators. In section 4, experimental methods and results are given. Experiments are constructed to compare methods in terms of computational times, accuracy and statistical significancy for discriminating motor/motor imagery tasks. For statistical significance, a new hierarchical clustering based procedure is proposed. Another test of significancy is implemented for selected electrode pairs. Finally in the Discussion and Conclusion sections, experiment results are evaluated in terms of study objectives and the thesis is

concluded with possible directions of future work.

CHAPTER 2

BACKGROUND

2.1. Functional Brain Imaging Techniques

A number of non-invasive brain imaging tools are used in neuroscience studies. These techniques can be divided into two groups. The first group includes techniques that directly aim to measure brain activity. EEG and MEG are placed in this group. The second group consists of methods like fMRI and PET that indirectly measure brain dynamics concerning cerebral vasculature (Mitra, 2007).

The earliest and the most commonly used brain imaging technique is electroencephalography (EEG). EEG measures the electrical voltage fluctuations of the brain by electrodes placed over scalp. Detailed information about EEG is given in the next section since the technique is used in this study.

The other method that directly measures brain activity is magnetoencephalography (MEG). Based on the fact that electric current generates a magnetic field, the first use of magnetic fields caused by bioelectric current sources was magnetocardiogram (Baule, 1963). After a few years, magnetic fields were used for detecting the alpha rhythms generated within the brain (Cohen, 1968). MEG measures the magnetic field produced by electric current emerging as a result of brain activity outside of the head. But the cortical surface of the brain consists of pyramidal neurons and these cells are predominantly directed perpendicularly to the scalp. Thus, because of the right hand rule, MEG receives the magnetic field predominantly from fissures and sulci. Additionally, magnetic fields received from the head are relatively much smaller than the outside sources such as the earth's natural magnetic field and magnetic field generated by powerline sources. Although various methods are used to compansate these outside magnetic sources, a magnetic shield or a magnetically shielded room is usually used for MEG (Laureys et al., 2002).

Another technique is functional magnetic resonance Imaging (fMRI). It uses the different features of atoms under a magnetic field. Three parameters are used for magnetic resonance imaging (MRI): proton density, longitudinal magnetization time and transverse

magnetization time. There are different methods used for MRI technique. For fMRI, mostly blood-oxygenation level dependent (BOLD) method is used. When a part of the brain is active, the oxygenated haemoglobin rate increases in this area and BOLD measures the oxygenated/deoxygenated haemoglobin rate with the help of different paramagnetic features of oxygenated and deoxygenated haemoglobin. Information is transmitted in milliseconds from one area to another in the brain, but the oxygen levels change more slowly. Because of that, the temporal resolution of this technique is poor. MRI was firstly used in 1973 and fMRI is invented in 1992 (Cabeza et al., 2006).

Positron emission tomography (PET) is one of the tissue autoradiography techniques. Before PET imaging, a positron emitting compound called radio tracer is injected in the vein of the patient. In a few minutes, this compound reaches the brain with the help of the bloodstream. After interacting with an electron, the tracer emits a gamma ray photon pair. These photons are detected by sensors placed around the patient in a ring and a three dimensional image of the brain is constructed. Although radioactive compounds had been used before, PET technology was developed as a brain imaging tool after computed tomography (CT) was invented (Ter-Pogossian et al., 1980). Temporal and spatial resolution of this technique is poorer than fMRI.

In addition to these methods, single-photon emission computerized tomography (SPECT), transcranial magnetic stimulation (TMS) and near-infrared spectroscopy (NIRS) are some of the techniques for functional brain imaging (Laureys et al., 2002).

The most significant features of brain imaging techniques are spatial and temporal resolutions. PET and fMRI tools provide better spatial resolution where EEG and MEG methods offer better time resolution (Figure 2.1). Especially while exploring the event related potentials (ERP), the temporal resolution is more important than spatial resolution. Because of that, EEG and MEG methods are used more for ERP activities. Although MEG has an advantage in temporal resolution, EEG is preferred more in studies because of the before mentioned direction problem of magnetic fields. Additionally, EEG is more compatible for a brain computer interface (BCI) application, since MEG devices are more expensive and have larger volumes because of the magnetic shielding. On the other hand, the lack of any volume conduction effects is an advantage of MEG. Because of that, some studies use MEG in addition to EEG (Nunes and Srinivasan, 2006).

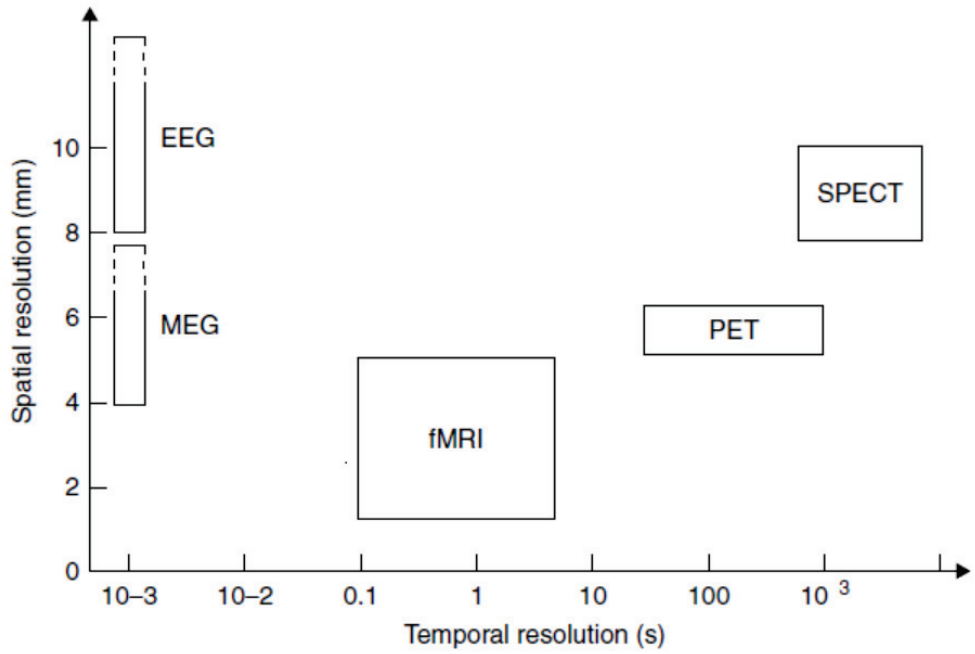


Figure 2.1. Temporal and spatial resolution comparison of brain imaging tools (Source: Laureys et al. (2002))

2.2. Electroencephalography (EEG)

The central nervous system (CNS) of humans consists of nerve and glia cells. Nerve cells (neurons) respond and transmit information by electrical impulses. Impulses generated by neurons are called action potentials (AP) and they last about 5-10 ms (Figure 2.2). A neuron generates an AP by transition of positive Na^+, K^+, Ca^+ and negative Cl^- ions from channels that exist on the axonal membrane. The interneuron information transmission is not as fast as the transmission of electricity within a wire: conduction velocity is in the interval of 1-100 m/sec (Sanei and Chambers, 2007).

Electrical activity of neurons causes an attenuated and noisy signal through the brain and skull. The attenuation is mostly caused by the skull. On the other hand, the noise is mostly because of the brain itself. Even though a neuron generates a potential between -60 mV and 10 mV, the signal peaks observed over scalp are measured between 10-100 uV.

Brain signals emerge in low frequency bands (100 Hertz effective bandwidth). These signals are analyzed in five major frequency bands: α (0.5-3.5 Hz), θ (3.5-7.5 Hz), β (7.5-12 Hz), δ (12-30 Hz) and γ (greater than 30 Hz). EEG measures those attenuated and noisy signals with low frequency.

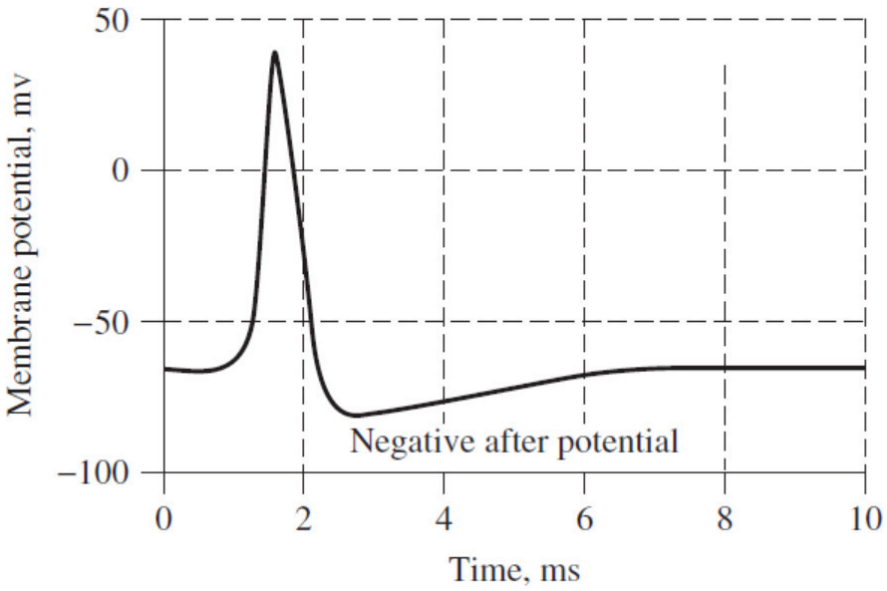


Figure 2.2. Action potential of a neuron (Source: Sanei and Chambers (2007))

The word electroencephalography means tracing electrical brain activity (electro-: electrical, encephalo-: related to brain, graphy: tracing). Cerebral activity was discovered long before, but the first recording of EEG was published by Hans Berger in 1929 (Berger, 1929). Early EEG signals were recorded by galvanometers. Recent EEG devices consist of electrodes, low noise amplifiers, ADCs and filters. In studies, EEG systems with different sampling rates, precision, electrode number and electrode placement systems are used. A conventional EEG system has a maximum 2000 of sample/sec rate since the signals have low frequencies. In addition, electrode numbers and placement system vary with the purpose of recording. Mostly used placement systems are 10/10 and 10/20 (Figure 2.3).

2.3. Brain-Computer Interface

Brain-Computer Interface (BCI) is a system that allows a user to control a device or provides input to the device by brain signals without requiring any movement. Even though BCI systems are typically used for disabled people such as those suffering from ALS (Amyotrophic Lateral Sclerosis), they can be also used by healthy people in various fields (Allison et al., 2007). A general BCI system framework is seen in Figure 2.4. Brain signals are received by the help of sensors from the user. Preprocessing operations

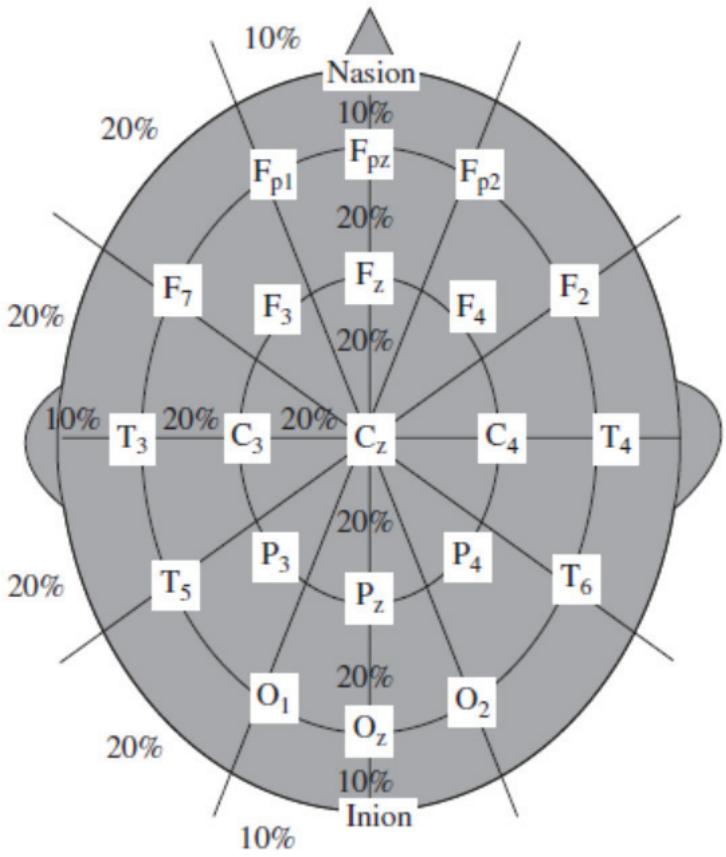


Figure 2.3. 10-20 Electrode placement of 21 electrodes (Source: Sanei and Chambers (2007))

such as filtering, up/down sampling, artifact removing etc. are applied on these amplified signals. After that, various features of preprocessed signals are extracted and used for classification. Finally, classification results are sent to the device in the form of commands. In addition, BCI systems include a user interface to help user control devices as in P300 speller example (Farwell and Donchin, 1988). (Bashashati et al., 2007) presented a review of signal processing algorithms on BCI applications with EEG signals.

Most commonly used technique to receive biosignal is EEG because of its low cost, advantage on temporal resolution and portability potential (Ranky and Adamovich, 2010). The majority of BCI studies with EEG are based on a specific paradigm on brain signals. These signals are event related potential (ERP), steady-state visual evoked potential (SSVEP) and event related synchronization/desynchronization (ERS/ERD). ERP signals emerge as a response of the brain to outside world stimuli as in the P300 speller procedure (Farwell and Donchin, 1988). Another approach, SSVEP is based on visual stimulation. A visual stimulation that is shown to the subject in a particular frequency

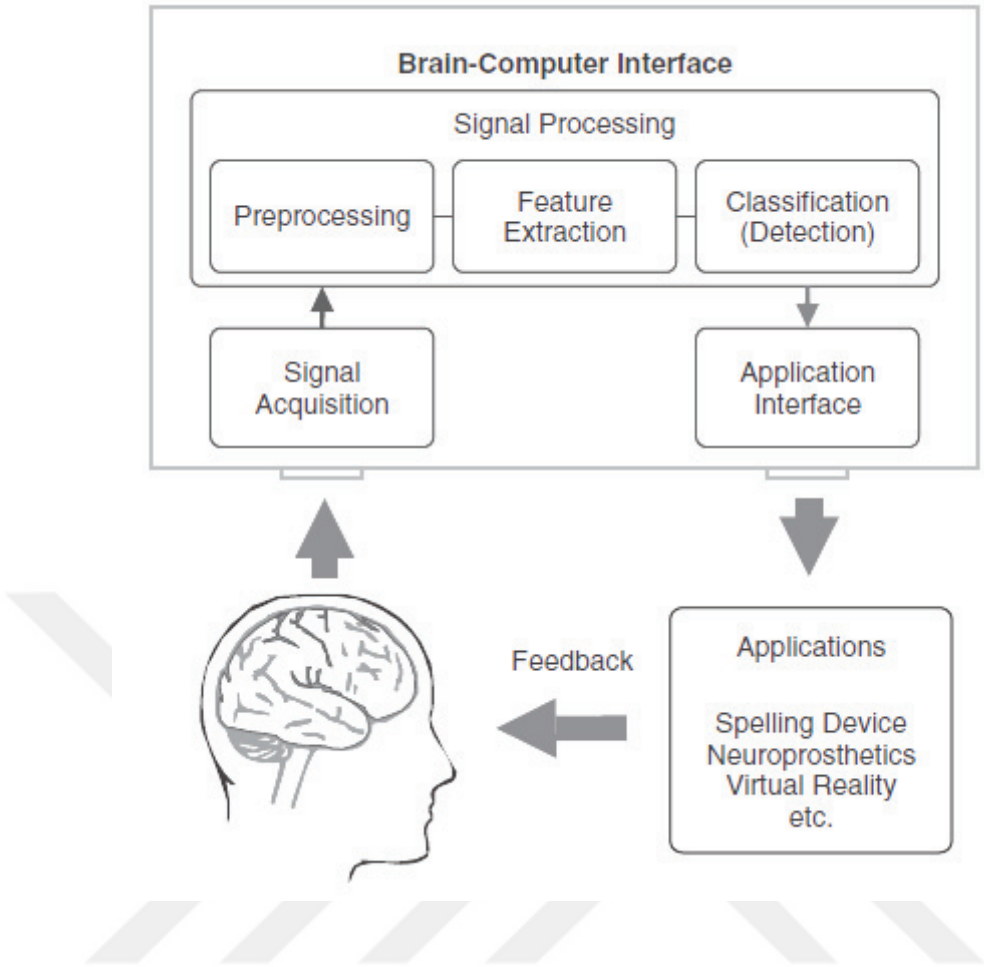


Figure 2.4. General BCI system framework (Source: Ahsan et al. (2009))

causes a signal at the same frequency in the users brain and in SSVEP based BCI systems stimuli with different frequencies help the user control the device (Zhu et al., 2010). And finally, ERS/ERD signals are used by detecting the synchronization and desynchronization patterns caused by different stimuli (Pfurtscheller and Neuper, 2006). Functional brain connectivity has recently received attention in BCI systems as a feature extraction option (Hamedi et al., 2016).

2.4. Functional Brain Connectivity

Functionality studies of the brain are carried out in two perspectives: functional segregation and functional integration. Functional segregation studies explore the functionally specialized areas for perceptual, motor, cognitive tasks etc. On the other hand, functional integration addresses the interaction between different specialized areas that

occurs under different conditions (Friston, 1994). Earlier neuroscience studies explored mostly the functional segregation or localization of activities. But there has been a shift to the functional integration topic in recent years (Friston, 2011).

Brain connectivity is categorized in three groups:

- Structural connectivity
- Functional connectivity
- Effective connectivity

Structural or anatomical connectivity looks for physical links between neurons or neuron groups. This connection type is not used for functional integration; generally, magnetic resonance imaging technologies are used to study structural connectivity. The second connectivity type, functional connectivity aims to find statistically correlated or dependent areas under certain actions. Finally, effective connectivity is concerned with the connection where an area triggers the activation of another area. The main difference of functional and effective connectivity is that effective connectivity requires causality along with a model for connectivity patterns, whereas functional connectivity is only interested in statistical relations as mentioned above (Sporns, 2010).

Brain connectivity studies have several purposes in general. One of them is to explore neurological disorders such as dyslexia (Odegard et al., 2009), schizophrenia (Cheung et al., 2008), autism (Belmonte et al., 2004) and epilepsy (van Diessen et al., 2013). Other studies aim to understand how brain actually works. (Varela et al., 2001) showed this curiosity with a question: “How does the brain orchestrate the symphony of emotions, perceptions, thoughts and actions that come together effortlessly from neural processes that are distributed across the brain?”. And recently, brain connectivity concept began to be explored in another area: brain computer interface (BCI) (Daly et al., 2012).

2.5. EEG and Brain Connectivity

EEG and fMRI are the two mostly preferred brain imaging techniques for functional connectivity studies. EEG is used more especially to detect relatively high speed communications between different brain areas since fMRI suffers from poor temporal resolution (Stam et al., 2007). For the analysis of functional connectivity with EEG, the statistical dependence of electrodes placed in different areas over the scalp is explored, though not directly of neuron groups.

Figure 2.5 illustrates the procedure for functional brain connectivity studies with EEG. Time series receieved from electrodes of EEG are used to compute a dependence matrix. This matrix includes a dependence value for each electrode pair. For the 64 channel dataset used in this study, 2016 dependence values are computed to form the dependence matrix. Various methods to measure dependency can be used to calculate this matrix. These methods will be explained further in the next section since this is the main topic of this thesis. Finally, this matrix is used to detect the connectivity patterns of the brain.

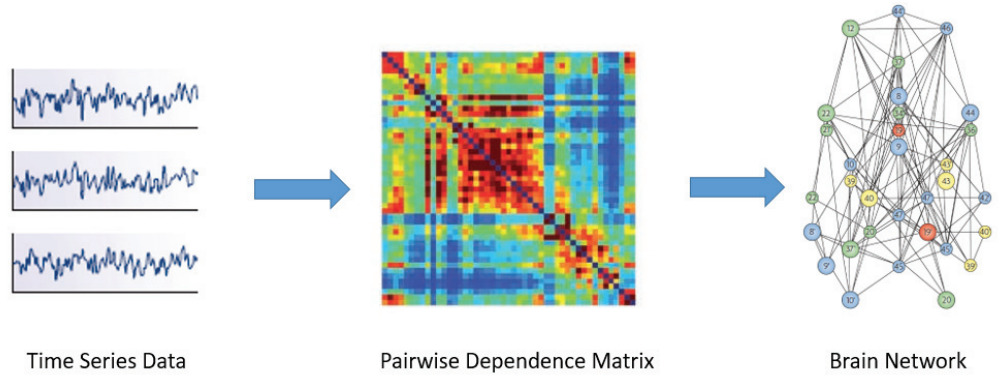


Figure 2.5. Functional connectivity procedure for EEG signals

Dependency of EEG time series may be observed with a time lag between electrodes because of inter-neuron communication delays. This causes a difficulty for measuring dependency since these delays are not known a priori. On the other hand, methods like phase locking value (PLV) take advantage of this time lag by analyzing the phase stability of time series as a measure of dependency (Rosenblum et al., 1996). Furthermore, despite the advantage of temporal resolution, brain connectivity studies with EEG suffer from volume conduction effects. This problem occurs because of the conductivity of the skull and the scalp (Peraza et al., 2012). EEG signals are not received from just neurons beneath the electrode but a linear mixing of neurons in a circle with a radius of about 6 cm (Thatcher et al., 2004). Although contribution of neurons weakens as the distance between electrodes grows, volume conduction problem causes information loss for connectivity studies. Because of that, several methods that claim to eliminate volume conduction were proposed such as phase lock index (PLI) (Stam et al., 2007) and imaginary part of coherence (iCOH) (Nolte et al., 2004).

CHAPTER 3

MATERIAL AND METHODS

In this chapter, we provide a description of EEG motor/motor imagery dataset used in the study and the preprocessing carried out on this dataset. After that, dependence measures are expressed in detail. In addition to theoretical background, information about the advantages and gaps of the measures are tried to be exposed in accordance with the features of EEG signals.

3.1. Dataset and Preprocessing

This section elaborates on the dataset used in the study as well as the various pre-processing steps involved in data preparation.

3.1.1. Dataset

In the study, an EEG motor movement/imagery dataset provided by the PhysioNet database was used (Goldberger et al., 2000). In the dataset, EEG data was recorded by a BCI2000 system (Schalk et al., 2004) with 160 Hz sample rate and the signal was received from a 64 channel cap with an international 10-10 system electrode placement (Figure 3.1).

There exist 14 experimental runs for each subject. The first two minute runs were recorded as baseline (the first run with eyes open, eyes closed for the other). The other runs have one of the tasks below and the tasks were performed multiple times in each run:

1. An indicator appears on the left or right side of the screen. The subject opens and closes the fist at the same side. The subject is requested to rest.

2. An indicator appears on the left or right side of the screen. The subject imagines to open and close the fist at the same side. The subject is requested to rest.

3. An indicator appears on the top or bottom of the screen. The subject opens and closes the both fist or both feet. The subject is requested to rest.

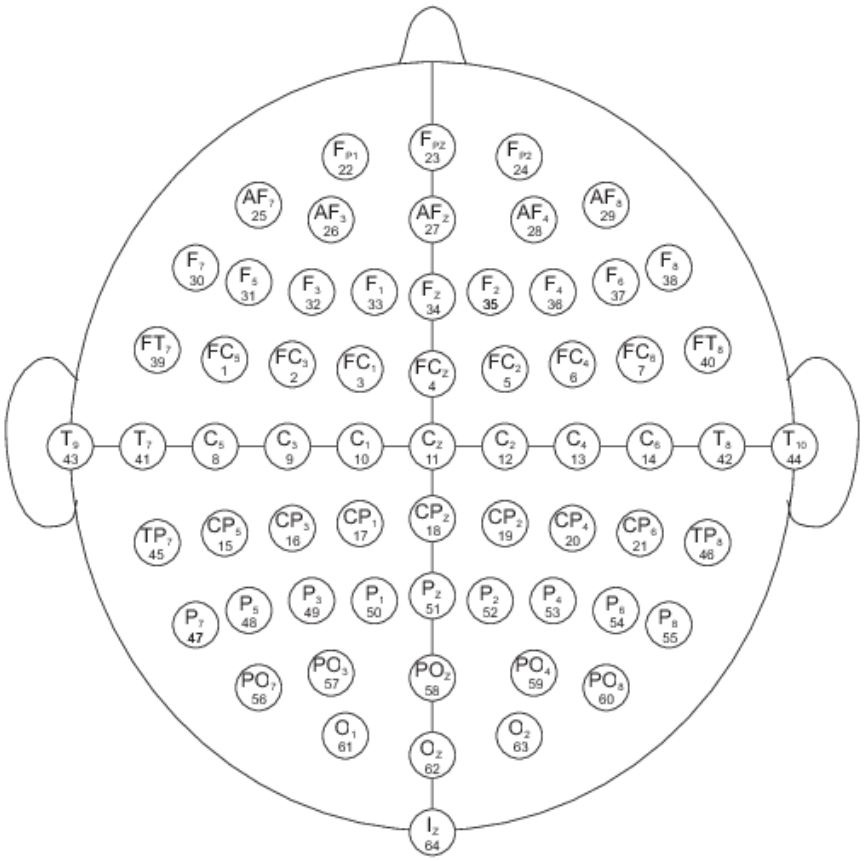


Figure 3.1. Electrode placement of PhysioNet dataset

4. An indicator appears on the top or bottom of the screen. The subject imagines to open and close the both fist or both feet. The subject is requested to rest.

Each experimental run was saved as `edf` formatted files (14 files per subject). In addition to each electrode data, one extra channel is used to hold the annotation, indicating the target codes and durations for each event. Table 3.1 summarizes event type and the target code in each experimental.

In total, 109 subjects participated in the experiment, but four subjects (subjects 88, 89, 92 and 100) were excluded from the study as their records deviate from the experimental paradigm described above. Furthermore, records of two additional subjects (subjects 43 and 104) were also excluded due to corrupt data (Loboda et al., 2014).

Table 3.1. Annotation and experimental run values for targets

Task	**Target Code**	**Run Index**
Left Fist	T1	3,7,11
Right Fist	T2	3,7,11
Imagine Left Fist	T1	4,8,12
Imagine Right Fist	T2	4,8,12
Both Fist	T1	5,9,13
Both Feet	T2	5,9,13
Imagine Both Fist	T1	6,10,14
Imagine Both Feet	T2	6,10,14
Rest	T0	all

3.1.2. Filter Design

The criteria below are followed for the design of the digital filters, compatible with the dataset:

- 0.5 - 40 Hz frequency band which includes the α, β, γ, δ and θ bands containing most information in EEG signal, should pass with maximum -6 dB attenuation at cutoff frequency.
- Group delay limit should be 50 ms for a real time BCI application filter.
- For the significance performance test accuracy of the dependence measures, group delay should be constant or almost constant for the [0.5,40] Hz interval.
- To avoid the power line effects, 60 Hz frequency (50 Hz for the most of the countries but the dataset is generated in USA) must be suppressed.

In accordance with these objectives, various infinite impulse response (IIR) and finite impulse response (FIR) filters were evaluated. IIR filters realize magnitude response more effectively even in low order cases. But it may cause a decline in dependence measure performance due to nonlinear phase response. On the other hand, even though FIR filters have linear phase, higher order filters are needed especially to implement narrow pass band and that means higher group delay (Oppenheim and Schafer, 2014).

In order to overcome the phase linearity problem of IIR, firstly a forward-backward filter design was implemented. But finally, because of the causality issue of this method,

an IIR and a FIR filter were used in a cascade structure to realize a band-pass filter with a pass-band of 0.5-40 Hz in the study.

3.1.2.1. Forward-Backward Filter Design

Forward-backward filtering aims to zero the phase response of any filter design by filtering the time-reversed signal after the first filtering process (Mitra and Kuo, 2006).

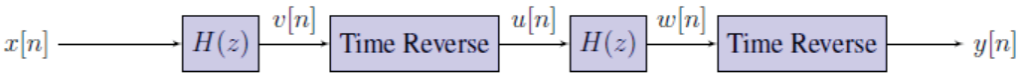


Figure 3.2. Forward-backward filtering block diagram

The block diagram of forward-backward filtering method is shown (Figure 3.2). It can be shown that the system only effects the magnitude of system (appendix B.1). Because of that, it is also called a *zero-phase filter*. For the forward-backward filtering, while MATLAB provides *filtfilt* function, some connectivity tools like *EEGLAB* offer an api of this method (Delorme and Makeig, 2004).

By following the method, a band pass filter design in interval of [0.5,40] Hz is aimed and target attenuation is limited to -3dB in cutoff frequency. Since theoretically the signal will be attenuated by the square of filter's magnitude response as a consequence of forward-backward filtering, a 14th order Butterworth filter that achieved -3dB attenuation at 0.2 Hz and 42 Hz frequency levels was used.

Frequency, phase and group delay response is shown (Figure 3.3) and in Figure 3.4, frequency, phase and group delay response of forward-backward filter constructed by cascaded and reversed linked Butterworth filters.Finally, to eliminate the effect of initial and end conditions of filtering, the signal is extrapolated with mirroring. Mirror length was chosen as the three times the filter order (Krauss et al., 1994).

Although the forward-backward filtering can meet all of the objectives listed above, it cannot be implemented in a real system since it is a non-causal filter. So a mix of IIR and FIR filter was designed and used for the preprocessing purpose.

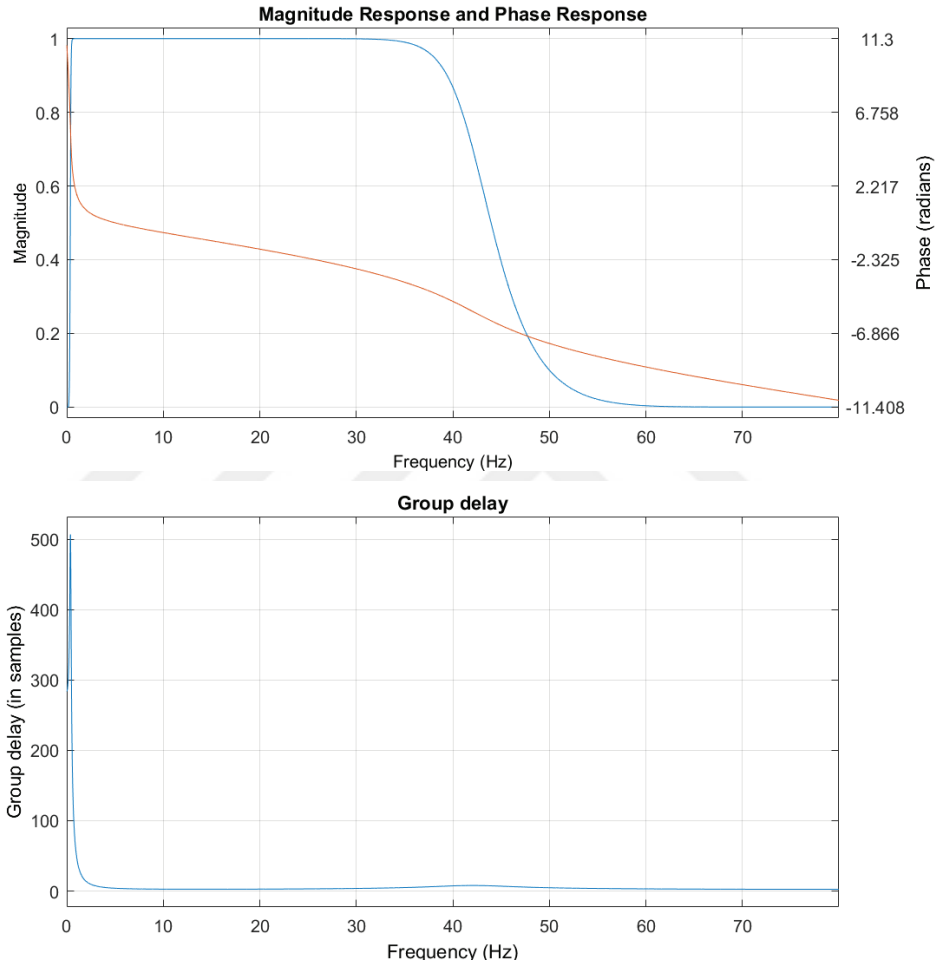


Figure 3.3. Response of the filter at the first stag

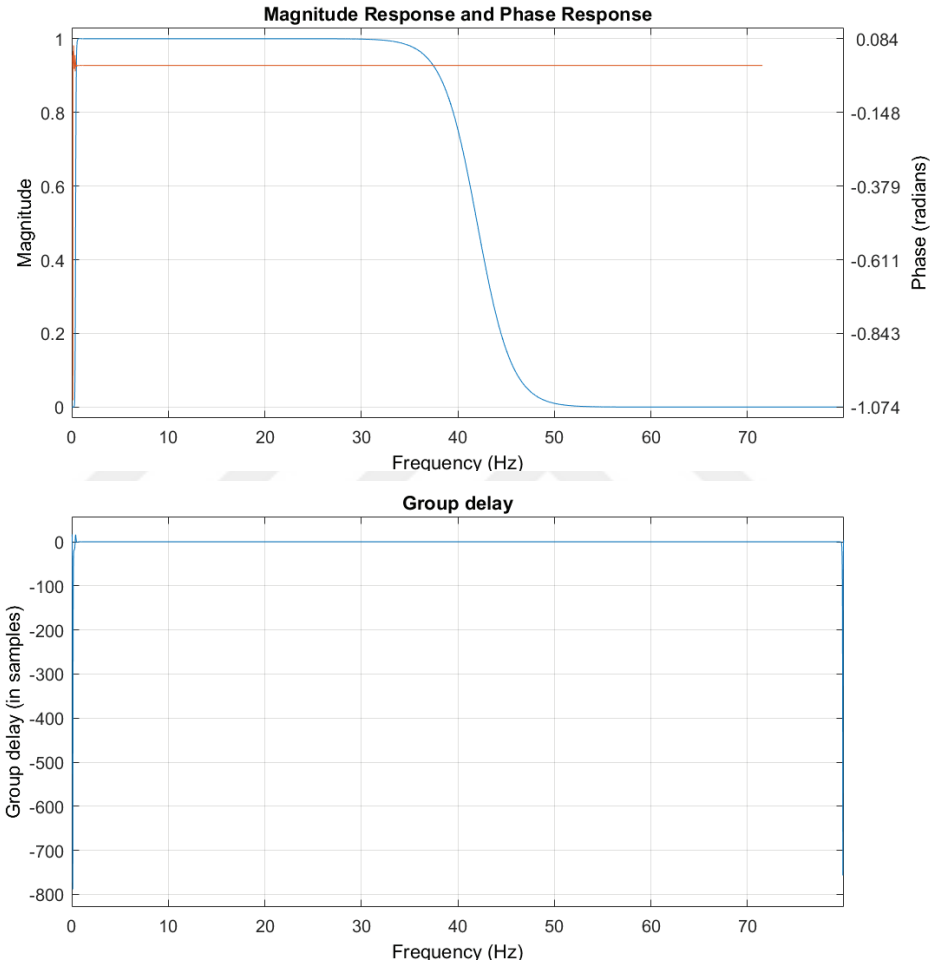


Figure 3.4. Response of the total forward-backward filtering

3.1.2.2. Butterworth IIR + Least Square FIR

A 16th order FIR filter which has a band pass between 35-45 Hz is designed by using least-squares method, when it is taken into consideration 160 Hz sampling frequency and 50 ms group delay constraint, to suppress the high frequencies. In addition, a first order Butterworth IIR filter with a -6dB 0.5 Hz cutoff frequency was used to eliminate the DC component.

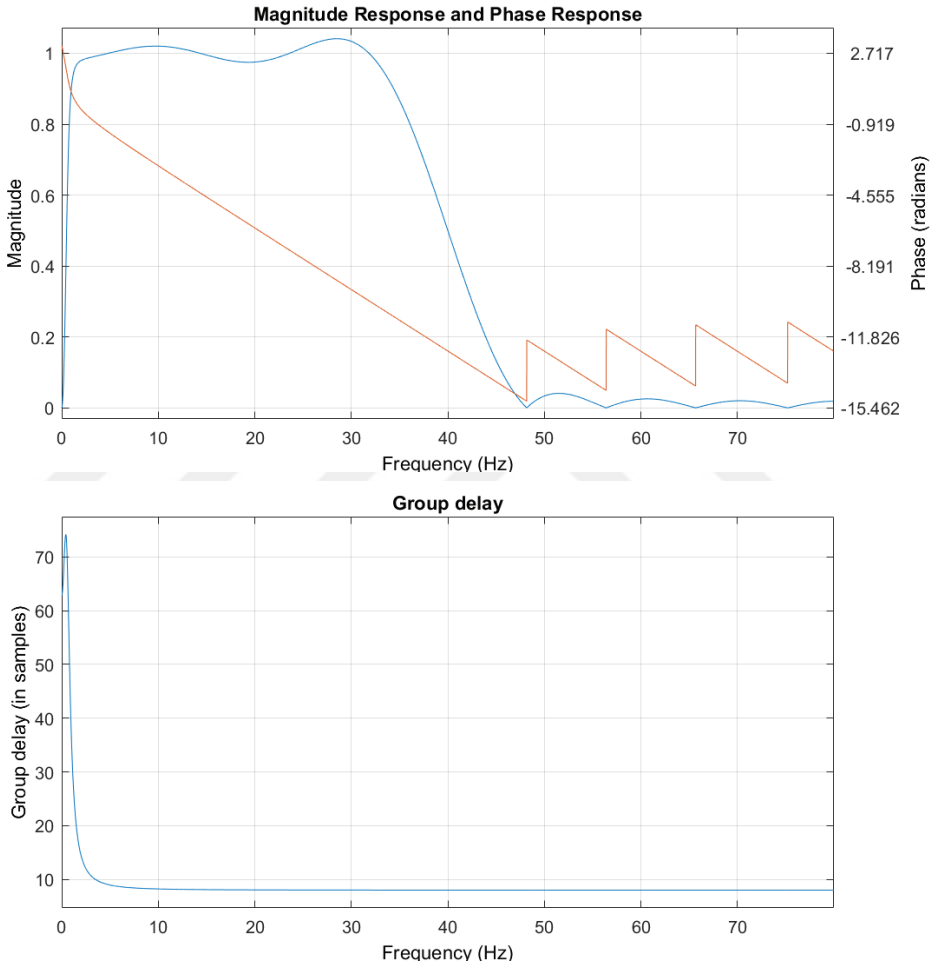


Figure 3.5. Response of the cascaded filter

The frequency, phase and group delay response of the cascaded filter is shown (Figure 3.5).

3.2. Dependence Measure Methods

A variety of methods have been proposed in the literature to evaluate the relationship between different EEG channels. These methods can be described in three categories:

- Correlation based methods.
- Phase synchronization based methods.
- Information based methods.

There is no doubt that strict categorization of these methods is not possible. Still, it is adopted here to help the description of the various methods. Finally, the method called *generalized synchronization* is explained independently outside of these categories.

3.2.1. Correlation Based Methods

One of the oldest and most basic way of measuring the relationship between two variables is *Pearson product correlation* (Pearson, 1920). The method simply calculates the proportion of covariance between the variables $(C_{xy}(X,Y))$ and the product of marginal standard deviation values $(\sigma_x(X), \sigma_y(Y))$:

$$r = \frac{C_{xy}(X,Y)}{\sigma_x(X)\sigma_y(Y)} \tag{3.1}$$

Pearson product correlation is easy to compute and used widely to measure the linear dependence of two variables. But it is sensitive to outliers and fails to evaluate nonlinear dependence. It should also be noted that uncorrelatedness does not guarantee independence.

Different methods are constructed over Pearson correlation to overcome the issues mentioned above: correlation in frequency domain, *spectral correlation method*; two widely used rank based methods *Kendall* τ and *Spearman* ρ, a kernel based canonical correlation method, *Schweiser-Wolff method* that measures the dependence by calculating the distance from the independence case in terms of copula functions. Details of these methods are provided below.

3.2.1.1. Spectral Correlation Coefficient

One of the most basic ways to evaluate the association between EEG channels is the Spectral Correlation Coefficient (SCC) method (Thatcher et al., 2004). SCC method measures the linear dependency between EEG channels' magnitude values in time or inter-trials over a frequency or a frequency interval. The method is implemented in two steps:

- Power spectra of two channels are computed for different epochs in time.
- Pearson product correlation value is calculated using the magnitude of a frequency interval for all epochs:

$$r(f) = \frac{\sum_{k=1}^{N}(X_k(f) - \overline{X}_k(f))(Y_k(f) - \overline{Y}_k(f))}{\sum_{k=1}^{N}(X_k(f) - \overline{X}_k(f)) \sum_{k=1}^{N}(Y_k(f) - \overline{Y}_k(f))} \tag{3.2}$$

In the expression above $X_k(f)$ and $Y_k(f)$ denotes the magnitude of power spectra of signals $x[n]$ and $y[n]$ at frequency f where k is the index of the epoch. SCC method has been used in brain signals for more than 50 years and it is still widely used (Adey et al., 1961). (Maby et al., 2004) examined the correlation of Auditory Evoked Potential (AEP) spectral contents by SCC method. (Swarnkar et al., 2007) computed SCC values for different frequency bands to view the inter-hemispheric synchronization during sleep apnea and EEG arousals.

3.2.1.2. Spearman's ρ and Kendall's τ

Spearman's ρ (Spearman, 2014) and Kendall's τ (Kendall, 1938) methods are two well known and widely used rank-based correlation measurement methods. Spearman's ρ is the correlation of ranks of the observations of the two variables. Let X and Y be two univariate variables with $x_1, x_2, ..., x_N$ and $y_1, y_2, ..., y_N$ their observations at N different time instants. Let also $r_x[n]$ and $r_y[n]$ be defined as the order indices or ranks of the respective observations. Spearman's ρ is then calculated as the correlation between r_x and r_y:

$$\rho = \frac{C(r_x, r_y)}{\sigma_{r_x}\sigma_{r_y}} \tag{3.3}$$

(Lachaux et al., 2003) used Spearman correlation in time-frequency space to find interaction between frequency bands of intra-cranial brain signals. Kendall's τ, on the other hand, is based on the concordance of sample pairs. Two samples can be stated as concordant if and only if $(x_i - x_j)$ and $(y_i - y_j)$ have the same sign. τ value is then calculated as the ratio of concordant pairs to all pair combinations:

$$\tau = \frac{N_c - N_d}{\frac{1}{2}N(N-1)} \tag{3.4}$$

In Eq.(3.4), N_c and N_d denote the concordant and discordant pair numbers where denominator is the number of pair combination $\binom{N}{2}$. When compared to product correlation, Kendall and Spearman methods are more robust to outliers. Moreover, they can detect even nonlinear dependencies when the variables are monotonically related to each other (Bonita et al., 2014). An example is given in Figure 3.6, where the variables are related through $Y = X^{10}$. Even though the variables are fully dependent, Pearson correlation is 0.66 due to the nonlinearity, where Kendall's τ and Spearman's ρ are both 1. But still, Kendall and Spearman correlations may give zero values even if there is a relationship between variables and they provide more erratic results than pearson correlation for different data under weak dependence even if they have the same relation degree.

(Bonita et al., 2014) compared Kendall rank correlation with Spearman, Pearson correlations and mutual information for the discrimination of *eyes open, eyes closed no task* states.

3.2.1.3. Schweizer-Wolff Measure of Dependence

Two rank based formulas described above, Spearman's ρ and Kendall's τ return zero when the variables are independent, but the reverse proposition is not true: ρ and τ may be equal to zero even if the variables are strongly dependent. Another rank based dependence measure, Schweiser-Wolff (SW) method is proposed to address this issue (Schweizer et al., 1981) based on copula functions of the underlying distributions rather than rank correlations.

A two dimensional copula is a function $C : [0,1]^2 \rightarrow [0,1]$, maps from $\mathbb{I}^2$ to $\mathbb{I}$, where $\mathbb{I}$ denotes the interval $[0,1]$ (Nelsen, 2007). Let X and Y be two univariate variables. Irrespective of the distributions of X and Y, we have the random variables $U = F_x(X)$ and $V = F_y(X)$ governed by standard uniform distributions. Then, the joint cumulative distributions of X and Y can be expressed as a copula function in terms of the

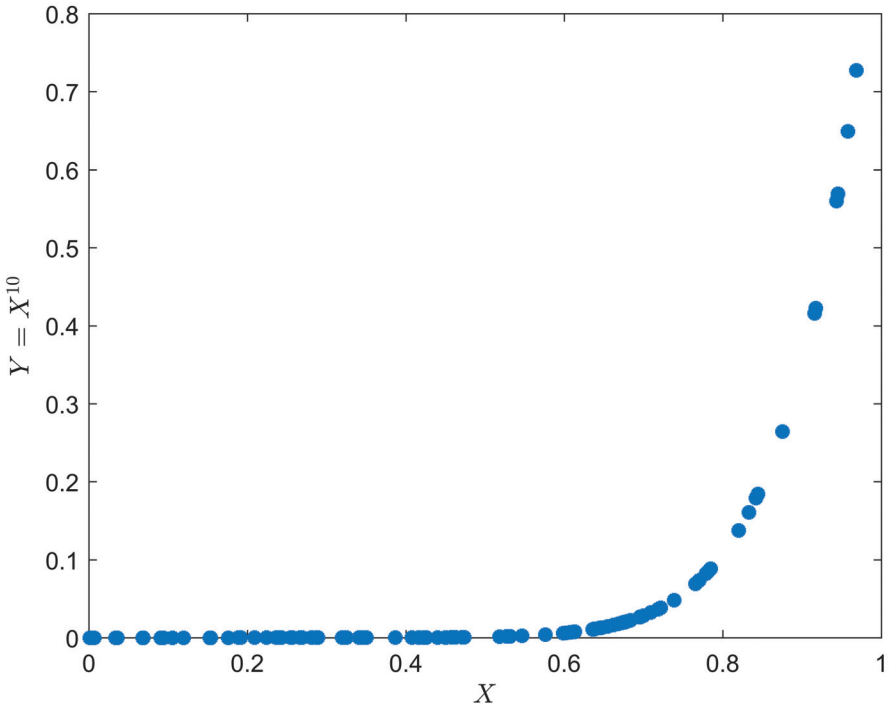


Figure 3.6. Fully nonlinear dependent variables

marginal cumulative distributions, U and V:

$$C(u,v) = F\left(F_x^{-1}(u), F_y^{-1}(v)\right) \tag{3.5}$$

When X and Y are independent, the copula function is simply calculated by the product of U and V, $C(u,v) = uv$. Because of that, it is called *product copula*, and represented as $\Pi(u,v)$.

Schweizer-Wolff method finds a dependence value by computing the normalized L_p norm of the difference between the product copula (independence case) and the two-dimensional copula functions estimated from the samples. The observed value when using the L_1 norm is termed σ, L_2 γ and L_∞ κ. Thus, σ, γ and κ are define as

$$\sigma(X,Y) = 12 \int_0^1 \int_0^1 |C(u,v) - uv| \, dudv \tag{3.6}$$

$$\gamma(X,Y) = \left(90 \int_0^1 \int_0^1 [C(u,v) - uv]^2 \, dudv\right)^{1/2} \tag{3.7}$$

$$\kappa(X,Y) = 4 \sup_{u,v\epsilon[0,1]} |C(u,v) - uv| \tag{3.8}$$

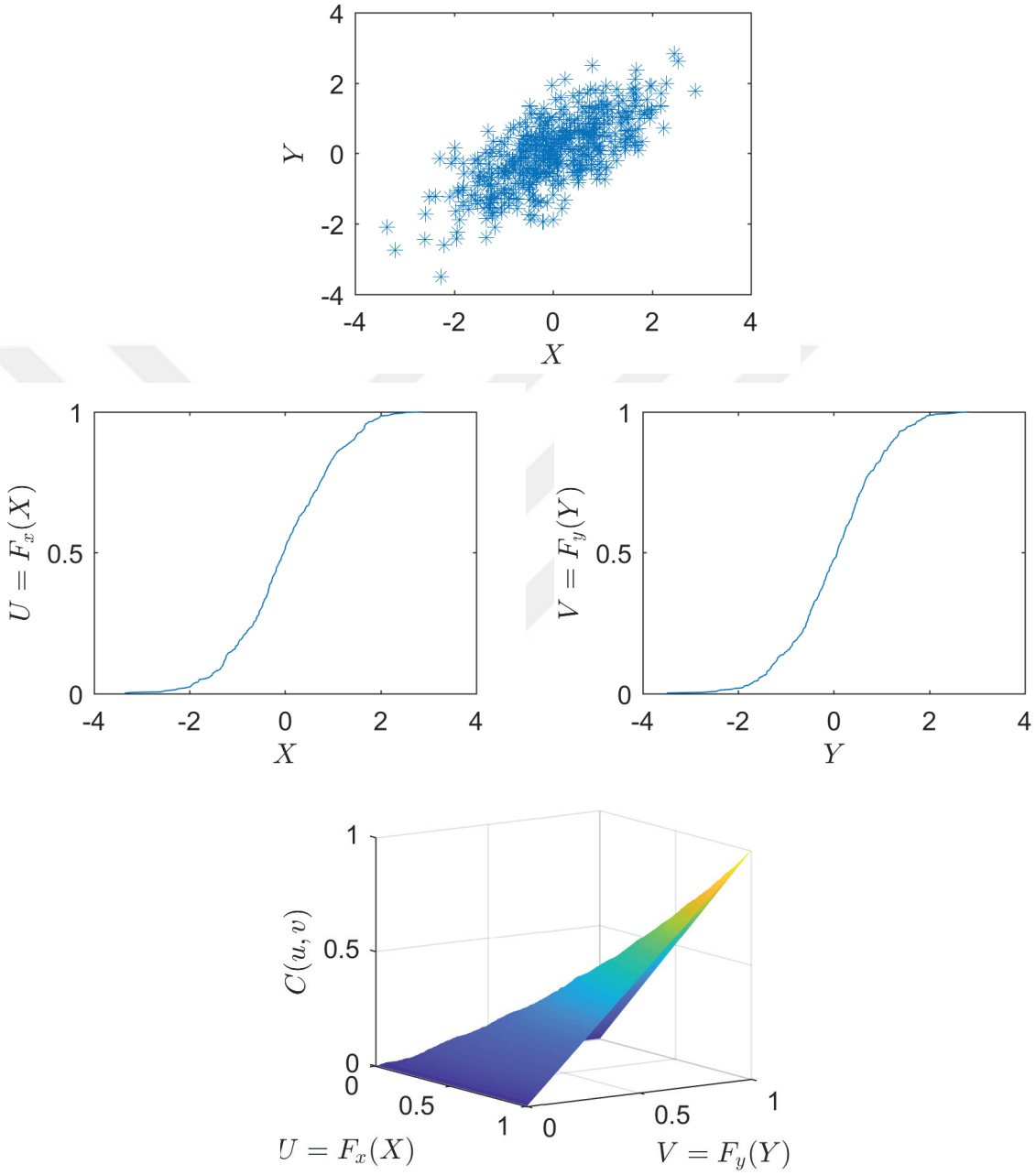


Figure 3.7. Example copula of two gaussian variables

Since the distributions of variables are not known, empirical copula estimation for the two-dimensional case is carried out using

$$C_N\left(\frac{i}{N}, \frac{j}{N}\right) = \frac{n_{ij}}{N} \tag{3.9}$$

In Eq.(3.9), N denotes number of samples and n_{ij} is defined as the number of samples (x_k, y_k) for which $x_k \leq x_i$ and $y_k \leq y_j$. Naturally, empirical estimation of the product copula equals $\frac{i}{N} \cdot \frac{j}{N}$. Therefore, the σ is calculated using empirical copula functions via

$$\sigma = \frac{12}{N^2 - 1} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| C_N\left(\frac{i}{N}, \frac{j}{N}\right) - \frac{i}{N} \cdot \frac{j}{N} \right| \tag{3.10}$$

SW method was used previously in an independent component analysis framework (SWICA) (Kirshner and Póczos, 2008). It is claimed to be robust to outliers, suffer less from noise and be relatively easy to implement.

3.2.1.4. Kernel Canonical Correlation Analysis

Even if two variables do not have a linear relation with each other and have a low Pearson correlation coefficient, this does not mean they are independent. As seen in Figure 3.8, there is a strong nonlinear dependence between two circular distributed variables even though Pearson correlation coefficient is almost zero. A kernel based correlation method is introduced (Akaho, 2006) in order to detect these kinds of nonlinear dependences.

Canonical Correlation Analysis (CCA) measures the dependence between two multivariate variables $\mathbf{x}$ and $\mathbf{y}$, where $\mathbf{x}\epsilon\mathbb{R}^{p_1}$ and $\mathbf{y}\epsilon\mathbb{R}^{p_2,}$ (Borga et al., 1997). In CCA procedure, two new variables $\mathbf{u} = \mathbf{a^T x}$ and $\mathbf{v} = \mathbf{b^T y}$ are generated under seperate linear transformations and a dependence measure $\rho(\mathbf{x}, \mathbf{y})$ is calculated by finding the linear transformation vectors $\mathbf{a}\epsilon\mathbb{R}^{p_1}$ and $\mathbf{b}\epsilon\mathbb{R}^{p_2}$ that maximizes the correlation between $\mathbf{u}$ and $\mathbf{v}$:

$$\begin{aligned} \rho(\mathbf{x}, \mathbf{y}) &= \max_{\mathbf{a},\mathbf{b}} corr(\mathbf{a^T x}, \mathbf{b^T y}) \\ &= \max_{\mathbf{a},\mathbf{b}} \frac{\mathbf{a^T C_{xy} b}}{\sqrt{\mathbf{a^T C_{xx} a b^T C_{yy} b}}} \end{aligned} \tag{3.11}$$

Taking the derivative of the correlation with respect to $\mathbf{a}$ and $\mathbf{b}$ and setting the result equal to zero provides

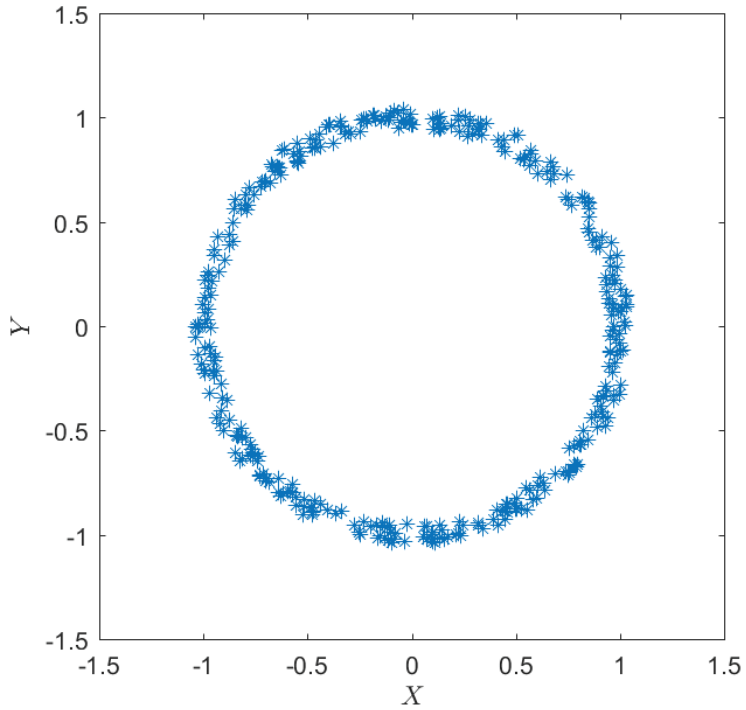


Figure 3.8. Two variables having nonlinear dependence in a circular distribution

$$\mathbf{C_{xy}b} = \frac{\mathbf{a^T C_{xy} b}}{\mathbf{a^T C_{xx} a}} \mathbf{C_{xx} a} \tag{3.12}$$

and

$$\mathbf{C_{yx}a} = \frac{\mathbf{b^T C_{yx} a}}{\mathbf{b^T C_{yy} b}} \mathbf{C_{yy} b} \tag{3.13}$$

Normalizing the vectors $\mathbf{u}$ and $\mathbf{v}$ by $\mathbf{a^T C_{xx} a} = 1$ and $\mathbf{b^T C_{yy} b} = 1$, the equations above can be expressed in the matrix form as

$$\begin{bmatrix} 0 & \mathbf{C_{xy}} \\ \mathbf{C_{yx}} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \rho \begin{bmatrix} \mathbf{C_{xx}} & 0 \\ 0 & \mathbf{C_{yy}} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \tag{3.14}$$

or equivalently as

$$\begin{bmatrix} \mathbf{C_{xx}} & \mathbf{C_{xy}} \\ \mathbf{C_{yx}} & \mathbf{C_{yy}} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = (1+\rho) \begin{bmatrix} \mathbf{C_{xx}} & 0 \\ 0 & \mathbf{C_{yy}} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \tag{3.15}$$

The problem thus turns out to be a generalized eigenvalue problem and the maximum eigenvalue is related to the correlation via $\lambda_{max} = 1 + \rho_{max}$.

Unlike CCA, KCCA method effectively maps the data in a nonlinear feature space to overcome the nonlinear dependence problem of Pearson correlation coefficient before finding the linear transformation. This is achieved by the so-called **kernel trick** described below.

If a kernel is an element of Reproducing Kernel Hilbert Space (RKHS), then it can be expressed as:

$$K(x, y) = \langle \boldsymbol{\phi}(x), \boldsymbol{\phi}(y) \rangle \tag{3.16}$$

in terms of some transformation $\boldsymbol{\phi} : \chi \to H$ maps the data from the observation space χ to the corresponding feature space, and $K : \chi \times \chi \to \mathbb{R}$ calculates the inner product in the transform space H denoted by $\langle ., . \rangle$. Thus, with the kernel, there is no need to know about the feature space and even infinite dimensional features can be dealt with via kernels. This is generally called "kernel trick". In order to kernelize Canonical Correlation Analysis, Eq.(3.11) is expressed over the transformed data in the transform space:

$$\rho(\mathbf{x}, \mathbf{y}) = \max_{\mathbf{a},\mathbf{b}} corr(\langle \boldsymbol{\phi}(x), \mathbf{a} \rangle, \langle \boldsymbol{\phi}(y), \mathbf{b} \rangle) \tag{3.17}$$

then, $\mathbf{a}$ and $\mathbf{b}$ will be redefined as $\mathbf{a} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\phi}(x^i)$ and $\mathbf{b} = \sum_{i=1}^{N} \beta_i \boldsymbol{\phi}(y^i)$. The covariance formula is also updated as

$$\begin{aligned} cov(\langle \boldsymbol{\phi}(x), \mathbf{a} \rangle, \langle \boldsymbol{\phi}(y), \mathbf{b} \rangle) &= \frac{1}{N} \sum_{k=1}^{N} \left\langle \boldsymbol{\phi}(x^k), \sum_{i=1}^{N} \alpha_i \boldsymbol{\phi}(x^i) \right\rangle, \left\langle \boldsymbol{\phi}(y^k), \sum_{j=1}^{N} \beta_j \boldsymbol{\phi}(y^j) \right\rangle \\ &= \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i K(x^k, x^i) K(y^k, y^j) \beta_j \\ &= \frac{1}{N} \alpha^T \mathbf{K_x} \mathbf{K_y} \beta \end{aligned} \tag{3.18}$$

where $\mathbf{K_x}$ and $\mathbf{K_y}$ are Gram matrices, it is possible to shape the covariance formula in a kernelized form without reading the actual transformation. Thus maximization can be carried out as

$$\rho(\mathbf{x}, \mathbf{y}) = \max_{\boldsymbol{\alpha},\boldsymbol{\beta} \epsilon \mathbb{R}^{\mathbb{N}}} \frac{\alpha^T \mathbf{K_x} \mathbf{K_y} \beta}{(\alpha^T {\mathbf{K_x}}^2 \alpha)^{\frac{1}{2}} (\beta^T {\mathbf{K_y}}^2 \beta)^{\frac{1}{2}}} \tag{3.19}$$

Finally, with respect to $\alpha, \beta \epsilon \mathbb{R}^N$ Eq.(3.19) can be rewritten just like Eq.(3.14) and the value is calculated by a similar eigenvalue decomposition operation:

$$\begin{bmatrix} 0 & \mathbf{K_x K_y} \\ \mathbf{K_y K_x} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \rho \begin{bmatrix} \mathbf{K_x K_x} & 0 \\ 0 & \mathbf{K_y K_y} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \tag{3.20}$$

When KCCA is applied on the circularly distributed variables similar to (Fukumizu et al., 2007) in the earlier example, after the kernelization process where a Gaussian Radial Basis Function (RBF) kernel defined as

$$K(x, y) = \exp\left(-\frac{1}{2\sigma^2}(x - y)^2\right) \tag{3.21}$$

The correlation is maximized at $\boldsymbol{\alpha} = -\mathbf{1}$ and $\boldsymbol{\beta} = \mathbf{1}$ to the level of 0.9810. For this specific example, Pearson correlation coefficient value for X and Y was 0.0004 indicating no correlation (Figure 3.9).

(Bach and Jordan, 2002) used KCCA method for Kernel Independent Component Analysis with another method called Kernel Generalized Variance. In Kernel ICA, they improved upon the KCCA method with a regularization process and followed an algorithm based on Incomplete Cholesky Decomposition. Three different kernels were used for KCCA in kernel ICA: namely the the classical polynomial kernel of order d defined by

$$K(x, y) = (r + sxy)^d \tag{3.22}$$

the $k - th$ Hermite kernel of order d defined by

$$K(x, y) = \sum_{k=0}^{d} \exp(-\frac{x^2}{2\sigma^2}) \exp(-\frac{y^2}{2\sigma^2}) \frac{h_k(x/\sigma)h_k(y/\sigma)}{2^k k!} \tag{3.23}$$

and the Gaussian RBF kernel shown in Eq.(3.21). For classical polynomial kernel r and s denote the coefficients, and for Hermite kernel definition $h_k(x)$ denotes the Hermite polynomial.

3.2.2. Phase Synchronization Based Methods

Phase synchronization refers to the stability in phase difference of coupled oscillatory systems (Thatcher et al., 2004). Two of the most used methods in EEG connectivity, *coherence* and *phase locking value (PLV)* are examined in this section.

Coherence is actually constructed on the product correlation concept but it is described here as it uses the phase information and it is easier to evaluate its similarity and difference to the PLV method. Following these two methods, *phase lag index(PLI)* and

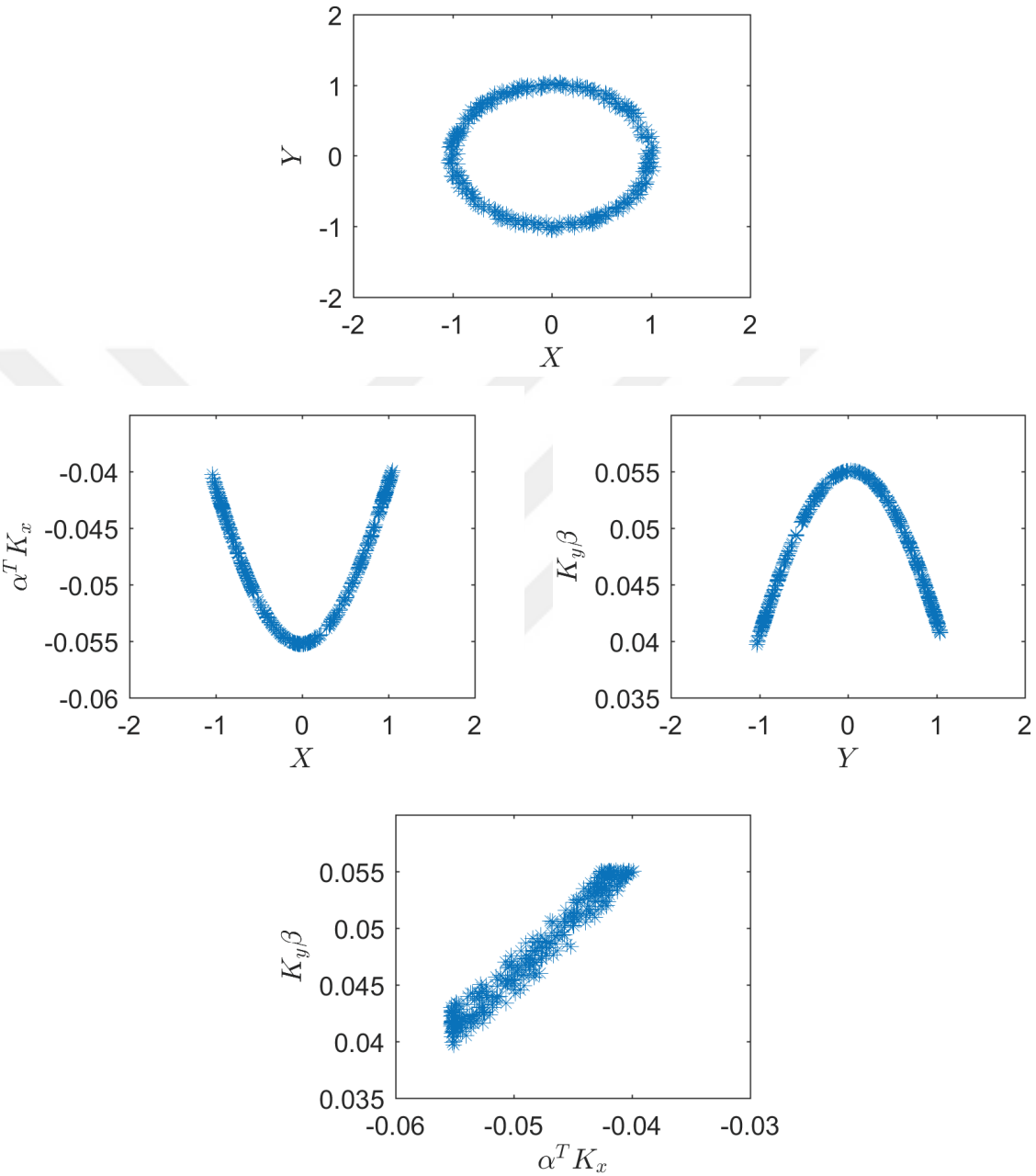


Figure 3.9. Applying KCCA on the circular distributed variables

imaginary part of coherence (iCOH) are also described as alternative methods proposed to avoid volume conduction effect.

3.2.2.1. Coherence

The spectral correlation coefficient method takes into account the magnitude information only and ignores possible phase relationships. On the other hand coherence method measures the linear relationship between two time series as a function of specific frequency by using both magnitude and phase. The method calculates the cross spectrum normalized by auto-spectral density function (Sakkalis, 2011) and it evaluates phase stability, defined as

$$r(f) = \frac{|\langle S_{XY}(f)\rangle|^2}{|\langle S_{XX}(f)\rangle|\,|\langle S_{YY}(f)\rangle|} \tag{3.24}$$

where $S_{XY}(f)$ denotes the cross spectrum of signals X and Y and $\langle . \rangle$ denotes the average value across time. Cross-spectrum of two signals can be computed either by performing the Fourier transform to the cross-correlation function or by multiplying the two signals in frequency domain as

$$S_{XY}(f) = X(f)Y^*(f) \tag{3.25}$$

In the case of absolute dependence, the coherence is 1 whereas a coherence value of 0 indicates random phase stability. The averages required for coherence are calculated over a number of time instants. In EEGLAB Toolbox, coherence measure is implemented using short time fourier transform and it is referred to as Event Related Linear Coherence (ERLCOH) (Delorme and Makeig, 2004). Generally, the coherence method can be applied to stationary signals while the use of STFT somewhat relaxes the assumption of stationarity (Sakkalis, 2011). On the downside, window length remains to be set as a parameter for the STFT operation. The window length parameter could lead to misleading results if selected inappropriately. (Lachaux et al., 2002) proposed using wavelets for coherence estimation and they called the method wavelet coherence (WC).

3.2.2.2. Phase Locking Value

Phase Locking Value or Phase Coherence is used to measure phase difference consistency of two signals in time. The method does not use amplitude information and differs from coherence method in this regard. It was proposed by (Rosenblum et al., 1996) to find phase synchronization between chaotic oscillators. They computed PLV over instantaneous phase value obtained using Hilbert Transformation detailed in appendix B.2, defined for a time-varying signal $x(t)$ as

$$\check{x}(t) = \frac{1}{\pi} PV \int_{\infty}^{-\infty} \frac{x(t')}{t - t'} dt' \tag{3.26}$$

where PV dentotes the Cauchy Principle Value. A new complex signal

$$H(t) = x(t) - \check{x}(t) \tag{3.27}$$

is formed by combining the original signal and the Hilbert transformed signal. The instantaneous phase $\phi(t)$ is then calculated easily by the angle of $H(t)$:

$$\phi(t) = \arctan\left(\frac{\check{x}(t)}{x(t)}\right) \tag{3.28}$$

For discrete time signals $x[n]$ and $y[n]$, the difference of phase values $\phi_x[n]$ and $\phi_y[n]$ are defined as

$$\Delta\phi[n] = |\phi_x[n] - \phi_y[n]| \tag{3.29}$$

The consistency of phase difference in time is computed as the magnitude of average value of the phase difference on the complex exponential unit circle via

$$PLV = \left| \frac{1}{N} \sum_{n=1}^{N} e^{j(\phi_x[n] - \phi_y[n])} \right| \tag{3.30}$$

In Figure 3.10, two example of instantaneous phase difference samples of time series and their average values are drawn on the complex exponential unit circle. The magnitude of consistent phase difference average is larger than the other one and it approaches to 1 as expected. A value of 0 means no phase coherency, while a value of 1 means total coherence.

Rosenblum applied PLV method to ECG signals with (Schäfer et al., 1999). After that, (Mormann et al., 2000) studied phase coherence on EEG records of epilepsy patients. (Lachaux et al., 1999) measured phase locking value in brain signals by estimating

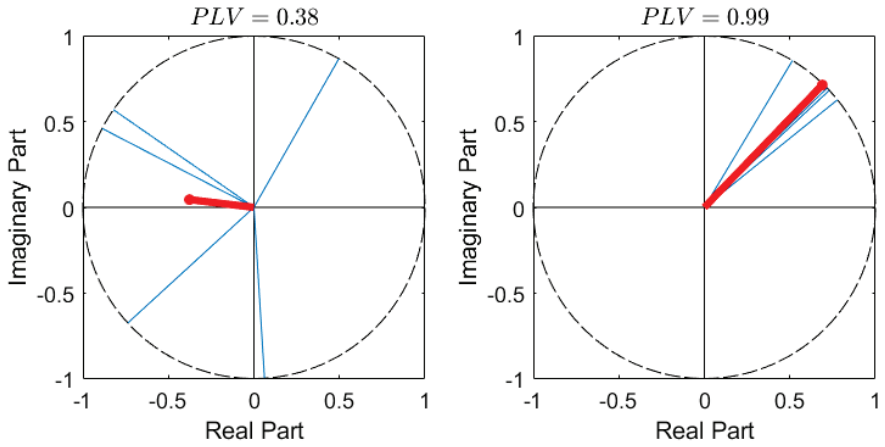


Figure 3.10. Illustirations for low and high phase stability

instantaneous phase value with a different method. They used wavelet transformation (detailed in appendix B.3) by simply computing the convolution of the signal with a complex Gabor wavelet defined by the product of a Gaussian and a sinusoidal signal at a frequency f, expressed as

$$\psi(t, f) = \exp^{\left(-\frac{t^2}{2\sigma_t^2}\right)} \exp^{j2\pi ft} \tag{3.31}$$

So the phase value $\phi_x(t)$ and $\phi_y(t)$ of the convolution is extracted for all time bins and PLV is again computed by the Eq. (3.30). Hilbert transformation and wavelet transformation have different advantages and thus different areas of use: Hilbert transformation is really good for narrow band signals but it may have some trouble for broadband signals and may require pre-filtering. On the other hand, wavelet transform is computed in a frequency centralized way and it is more capable at computing phase coherence around a given frequency and for the variance of the Gabor Wavelet.

3.2.2.3. Phase Lag Index

Phase difference between two signals will either be in the interval of $-\pi < \Delta\phi < 0$ or $0 < \Delta\phi < \pi$. The difference being in the positive or negative side consistently in a time period can provide information about phase synchronization for electrode pairs. To this end, the Phase Lag Index (PLI) can be used as a way to evaluate the asymmetry in the distribution of phase differences between two time series (Stam et al., 2007), defined as

$$PLI = |\langle sign\left[\Delta\Phi\left(t\right)\right]\rangle| \tag{3.32}$$

$\Delta\Phi\left(t\right)$ denotes the phase difference function of the two variables over time. The main objective of PLI method is to estimate phase synchronization while avoiding dependencies caused by the volume conduction effect. Since volume conduction causes dependencies that will have almost zero lag and zero phase difference, the phase difference value distribution will be centralized around 0 in modulo π and PLI will be close to zero.

Although PLI is a good way to measure connectivity while avoiding volume conduction, it can miss possible zero lag dependencies that have phase differences close to zero. It is defined as conservative since it ignores zero lag associations (Demuru et al., 2013). Phase difference values can be calculated using either Hilbert Transformation or Wavelet Transformation as in the calculation of PLV.

3.2.2.4. Imaginary Part of Coherence (iCOH)

As explained in the Background, one of the features that makes EEG hard to process is the volume conduction effect: One source in the brain can effect multiple electrodes with different amplitudes. The basic assumption about the effect is that volume conduction causes zero-lag interaction between electrodes. Let $x(t)$ and $y(t)$ be two time series affected by L independent sources, with the magnitude values of signals interpreted as weighted sums of sources as

$$X(f) = \sum_{k=1}^{L} a_k S_k(f) \tag{3.33}$$

$$Y(f) = \sum_{k=1}^{L} b_k S_k(f) \tag{3.34}$$

Then the cross-spectrum of the two time series is:

$$S_X Y(f) =< X(f)Y^*(f) >= \sum_{k=1}^{L} a_k b_k |S_k(f)|^2 \tag{3.35}$$

Because of the zero-lag assumption of volume conduction, as seen in Eq.(3.35), the cross-spectrum is real valued. If the coherence is defined as the sum of imaginary and real parts of coherence as $COH(f) = \Re(COH(f)) + \Im(COH(f))$, this means that all the volume conduction artifacts will add up to the real part and the imaginary part of

coherence will be insensitive to volume conduction effects. Thus, the imaginary part of coherence (iCOH) is introduced as a new interaction measure (Nolte et al., 2004):

$$iCOH = \Im(COH(f)) \tag{3.36}$$

Although iCOH addresses to overcome the volume conduction problem, it comes with various other issues. Firstly, neglecting real part means neglecting possible zero-lagged interactions. Secondly, iCOH is not normalized whereas the original coherence method gets value within the interval $[0, 1]$. And finally, it may fail to detect even the strong interactions if the real part is relatively large. Even though some methods have been proposed in the literature like dividing the square of iCOH with $(1 - \Im(COH(f)^2)$ (Pascual-Marqui et al., 2011) or using Hilbert transformation of $S_{XY}(f)$ (Bornot et al., 2018) to solve the normalization problem, the basic iCOH method is used in this study as is because of its widespread use.

3.2.3. Information Theoretic Methods

Information theory begins with the 1948 paper of Claude Shannon who was a mathematician in Bell Labs (Shannon, 1948). He is the first person to introduce the entropy concept to measure the amount of information associated with a random variable X via

$$H(X) = \int_{-\infty}^{\infty} f(x) \log f(x) dx \tag{3.37}$$

The Shannon entropy simply calculates the expected value of the logarithm of the probability density function $f(x)$. Furthermore, he also introduced the mutual information as a method to determine the rate of transmission on a communication channel and defined it as the difference between the sum of marginal entropies and the joint entropy of two random variables X and Y, as

$$I(X, Y) = H(X) + H(Y) - H(X, Y) \tag{3.38}$$

Note that the mutual information of X and Y can also be expressed as the distance between the joint entropy and the product of marginal entropies:

$$I(X, Y) = \int \int f(x, y) \log \frac{f(x, y)}{f(x) f(y)} \tag{3.39}$$

Mutual information is a useful tool to measure the dependence between random variables in many engineering areas like signal processing, machine learning and communication. One thing that makes the method preferable is that it does not only detect the dependence caused by linear covariance but the nonlinear ones as well.

After Shannon, Alfred Renyi proposed an α parameter generalization of entropy (Rényi, 1961):

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left[\int f_x^\alpha(x) dx \right] \tag{3.40}$$

As $\alpha \to 1$, Eq.(3.40) approaches the Shannon entropy. Finally, Renyi information is defined as

$$I_\alpha(X,Y) = \frac{1}{1-\alpha} \log \int f^\alpha(x,y) \left[f_x(x) f_y(y) \right]^{(1-\alpha)} dxdy \tag{3.41}$$

Information theoretic methods are highly preferred measures to study connectivity but both Renyi and Shannon information methods are hard to compute since the estimation of probability density function is a challenge in and of itself for continuous variables. Some widely used methods *(Kraskov, kernel generalized variance (KGV) and Rank Based Graph Optimization(REGO)* as information estimators are described below. In addition, two new estimation methods *(unit vector parametrization and data fitting based mutual information method)* are proposed as novel information theoretic measures for connectivity.

3.2.3.1. Kraskov Method

Kraskov's method (Kraskov et al., 2004) is based on the Kozachenko-Leonenko method (Kozachenko and Leonenko, 1987), which computes Shannon entropy with k-nearest neighborhood. More specifically, Kozachenko-Leonenko method calculates an entropy over the mean distances of the k-th nearest neighbor to each sample point x_i, $i = 1, 2, ..., N$:

$$\hat{H}(X) = -\phi(k) + \phi(N) + \log c_d + \frac{d}{N} \sum_{i=1}^{N} \log \epsilon(i) \tag{3.42}$$

In the above expression, N is the number of total samples, $\phi(x)$ the digamma function, d the random variable dimension, c_d the volume of the d-dimensional unit cube (area for 2-dimensions) and $\epsilon(i)$, is twice the distance of the k-th nearest neighbor to the sample x_i.

(Kraskov et al., 2004) have identified two similar methods of computing information through this entropy estimation formula. In this study, the following expression is used for comparison, as preferred later in the mutual-information-based least dependent component analysis (MILCA) method (Stögbauer et al., 2004):

$$\hat{I}(X,Y) = \phi(k) - \frac{1}{k} - \langle \phi(n_x) + \phi(n_y) \rangle + \phi(N) \tag{3.43}$$

In Eq.(3.43), $n_x(i)$ and $n_y(i)$ are defined as the number of elements that fit the conditions:

$$\|x_i - x_j\| \leq \epsilon_x(i)/2$$

and

$$\|y_i - y_j\| \leq \epsilon_y(i)/2$$

where $\epsilon_x(i)$ and $\epsilon_y(i)$ are the minimum edge lengths of the rectangle that includes the k-th nearest neighbor, and $\langle . \rangle$ is the average operator over all samples indexed by $i = 1, 2, ..., N$. k parameter was chosen as 1.

3.2.3.2. Unit Vector Parametrization

As already described above, it is enough to have marginal and joint entropy values to estimate the amount of mutual information of dependent variables. There are various methods for estimating marginal entropy in the literature (Wieczorkowski and Grzegorzewski, 1999). Among those methods, Vasicek entropy (Vasicek, 1976) that uses sample distances like Kozachenko-Leonenko entropy estimation (Kozachenko and Leonenko, 1987), was found to be fast and accurate (Wieczorkowski and Grzegorzewski, 1999). For any X univariate random variable, let $x_1, x_2, ..., x_n$ be the samples and $x_{(1)}, x_{(2)}, ..., x_{(n)}$ be the ordered values. Vasicek entropy calculates

$$V_{m,n} = \frac{1}{n} \sum_{i=1}^{n} \log \left(\frac{n}{2m} \left(x_{(i+m)} - x_{(i-m)} \right) \right) \tag{3.44}$$

As n and m tend to infinity subject to the condition $m/n \to 0$, $V_{m,n}$ value converges to true entropy value (Vasicek, 1976). To provide a more consistent estimator, an updated version of equation is recommended, specifying

$$H_V(X) = V_{m,n} - \log(n) + \log(2m) - \left(1 - \frac{2m}{n}\right) \phi(2m) + \phi(n+1) - \frac{2}{n} \sum_{i=1}^{m} \phi(i+m-1) \tag{3.45}$$

In Eq.(3.45), $\phi(x)$ is defined as the digamma function similar to the Kraskov method. $H_V(X)$ is also called as m_n-spacing estimate in some studies. The method is also used in an ICA algorithm called robust accurate direct ICA algorithm (RADICAL) (Learned-Miller and John III, 2003). Another method that uses a segmentation algorithm based on Voronoi regions and Delaunay cells also uses Vasicek entropy for joint entropy estimation (Miller, 2003), but it is a complex and time consuming solution. As an alternative, the unit vector parametrization method aims to offer a fast convergence to joint entropy and thus mutual information, using a property of entropy under linear transformation (Çağdaş and Karaçali, 2018). Suppose that random variables U and V are obtained by a linear transformation with a transformation matrix $\mathbf{A}$ applied on variables X and Y via

$$\begin{bmatrix} U \\ V \end{bmatrix} = A \begin{bmatrix} X \\ Y \end{bmatrix}$$

In this case, joint entropies, $H(U,V)$ and $H(X,Y)$ are related to each other through the expression

$$H(U,V) = H(X,Y) + \log|\det(A)| \tag{3.46}$$

Another property of entropy is that the joint entropy of variables is always smaller than the sum of the marginal entropy values:

$$H(U,V) \leq H(U) + H(V) \tag{3.47}$$

From Eq.(3.46) and Eq.(3.50), it is easy to infer the following relation between the joint entropy $H(X,Y)$ and the marginal entropies of variables U,V, obtained under a linear transformation:

$$H(X,Y) \leq H(U) + H(V) - \log|\det(A)| \tag{3.48}$$

The above expression means that the smaller the value of the right hand side of the equation, the more the value converges to the true joint entropy value. Minimum value point of these transformation matrices space can therefore be used as an estimate of joint entropy:

$$\hat{H}(X,Y) = \min_{A}(H_V(U) + H_V(V) - \log|\det(A)|) \tag{3.49}$$

Similarly, the above expression can be used in the estimation of mutual information as well together with Vasicek entropy estimation used to calculate the marginal entropies as

$$\hat{I}(X,Y) = H_V(X) + H_V(Y) - \hat{H}(X,Y) \tag{3.50}$$

In this method, search of the matrix A that minimizes the joint entropy is a major task. For this study, A matrices are generated by cosine and sine pairs like rotation matrices:

$$A(\theta_1, \theta_2) = \begin{bmatrix} \cos(\theta_1) & \sin(\theta_1) \\ \cos(\theta_2) & \sin(\theta_2) \end{bmatrix} \tag{3.51}$$

parametrized by angles $\theta_1, \theta_2 \epsilon [-\pi, \pi)$, and the matrix that computes minimum joint entropy is searched in this matrix sub-cluster with respect to θ_1, θ_2.

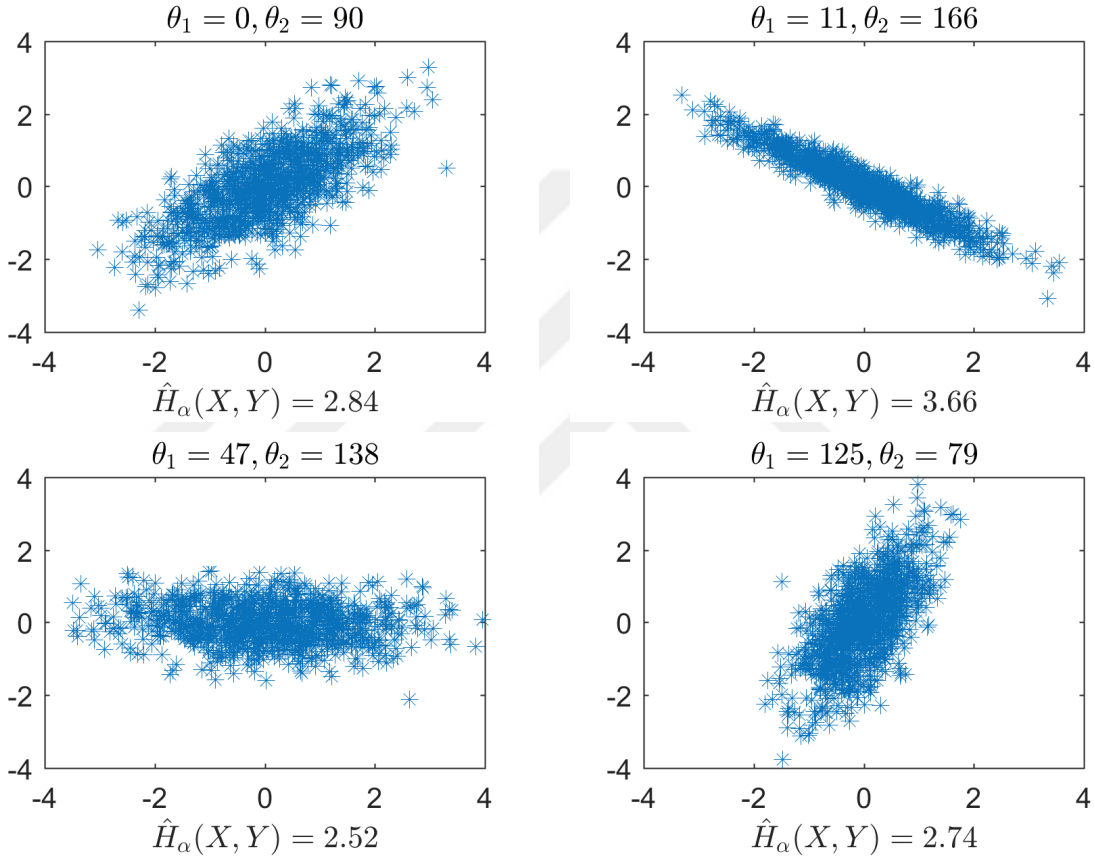


Figure 3.11. $\hat{H}_\alpha(X,Y) = H(U) + H(V) - \log|\det(A)|$ calculations for different θ pairs

In Figure 3.11, different linear transformation matrices are applied on jointly Gaussian random variables to estimate the joint entropy value. The case for $\theta_1 = 0$ and $\theta_2 = 90$ gives the original data. It is seen that the case for $\theta_1 = 47$ and $\theta_2 = 138$ has the lowest calculated value, and it converges to the theoretical value of the true entropy $H(X,Y) = \frac{1}{2}\log\left((2\pi\exp)^2|\Sigma|\right) = 2.5012$.

3.2.3.3. Data Fitting Based Mutual Information Calculation

Even though unit vector parametrization method provides acceptable results for mutual information estimation between linear dependent variables, it is ill-equipped for nonlinearly dependent variables. Data fitting based method can handle these kind of variables by trying to remove the conditionalities that can be observed between them. For this method, a conditional entropy based expression of mutual information is used, in accordance with Eq. (3.38):

$$I(X,Y) = H(Y) - H(Y|X) \tag{3.52}$$

In the independence of X and Y case, the equality $H(Y|X) = H(Y)$ emerges. In this sense, in order to make Y as independent as possible from X, an option is to remove the conditional expected value and conditional variance in relation to X from the variable Y. Note that while these efforts may not provide absolute independence hidden in higher order moments, it may help us reduce the conditional entropy sufficiently to allow an accurate assessment of the mutual information. This produces an alternative random variable Y' defined by

$$Y^{'} = \frac{Y - m(x))}{\sigma(x)} \tag{3.53}$$

In the expression above, $m(x)$ and $\sigma(y)$ are defined as $E[Y|X = x]$ and $\sqrt{Var(Y|X)}$, respectively. Thus an alternative approach for mutual information estimation can be expressed using two properties of conditional entropy:

1. If X and Y are two random variables, translating Y for any value of X does not change the conditional entropy value:

$$H(Y + c|X = x) = H(Y|X = x) \tag{3.54}$$

2. Given that $X = x$, multiplying Y by a constant value causes a change on the value of conditional entropy by the logarithm of the multiplication factor:

$$H(aY|X = x) = H(Y|X = x) + \log(a) \tag{3.55}$$

When these two properties -the accuracy of which can be easily proven- are considered, a difference of expected value of logarithm of conditional standard deviation between the conditional entropy of $Y^{'}$ and conditional entropy of Y can be obtained by

$$
\begin{aligned}
H(Y|X) &= H(Y - m(x)|X) \\
&= H\left(\frac{Y - m(x)}{\sigma(x)}|X\right) + E[\log \sigma(x)] \\
&= H(Y'|X) + E[\log \sigma(x)]
\end{aligned}
\tag{3.56}
$$

When we also consider that $H(Y^{'}|X) \approx H(Y^{'})$ neglecting higher-order dependencies, mutual information of X and Y can be estimated by the help of Vasicek Entropy estimation method as

$$
\hat{I}(X, Y) = H_v(Y) - H_v(Y') - E[\log \sigma(x)] \tag{3.57}
$$

The process of removing the mean and variance conditionality is illustrated (Figure 3.12). For this example, at the end of these operations an exact independence between variables is achieved.

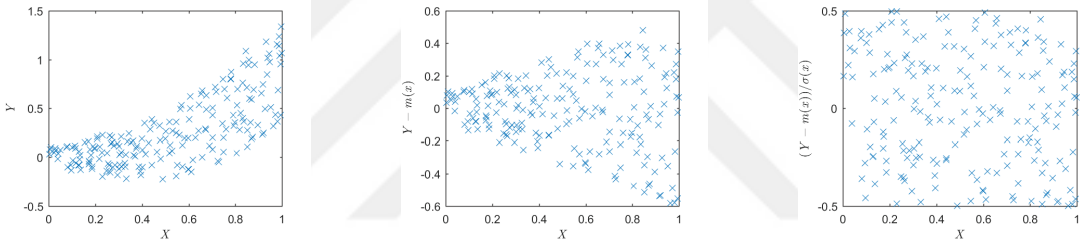


Figure 3.12. Removing the mean and variance conditionality from Y

When Eq.(3.57) is examined, estimation of the mutual information is reduced to the estimation of the conditional variance estimation $\sigma(x)$. In this study, for variance estimation, running mean smoothing and polynomial regression methods are used (Bishop, 2006), (Hastie and Tibshirani, 1990). These methods are explained briefly in appendix B.4 with the procedure followed for choosing the optimum polynomial degree parameter of the regression and the window length parameter of the smoother.

3.2.3.4. Kernel Generalized Variance

As mentioned in KCCA measure, one of the methods used as contrast function in Kernel ICA is Kernel Generalized Variance (KGV) (Bach and Jordan, 2002). KGV uses the relation of CCA and mutual information of Gaussian variables.

Let X and Y be two Gaussian variables that take values in $\mathbb{R}^{p_1}$ and $\mathbb{R}^{p_2}$ respectively. Then, the mutual information of these variables is related to the correlation coefficient via

$$I(X,Y) = -\frac{1}{2}\log\frac{\det(C)}{\det(C_{xx})\det(C_{yy})} \tag{3.58}$$

The expression inside the logarithm is called the generalized variance. Given the expression

$$\begin{bmatrix} \mathbf{C_{xx}} & \mathbf{C_{xy}} \\ \mathbf{C_{yx}} & \mathbf{C_{yy}} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = (1+\rho) \begin{bmatrix} \mathbf{C_{xx}} & 0 \\ 0 & \mathbf{C_{yy}} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \tag{3.59}$$

for CCA, it is seen that the equation is in the form of $C\xi = \lambda D\xi$ with $\lambda = 1+\rho$. Then the mutual information can be expressed as

$$\begin{aligned} I(X,Y) &= -\frac{1}{2}\log\frac{\det C}{\det D} \\ &= -\frac{1}{2}\log\prod_{i=1}^{p_1+p_2}(1+\rho_i) = -\frac{1}{2}\sum_{i=1}^{p_1+p_2}\log(1+\rho_i) \end{aligned} \tag{3.60}$$

Kernelized CCA which can also be expressed in the form of $K_K \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \lambda D_K \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, and this provides the so-called kernel generalized variance:

$$I(X,Y) = -\frac{1}{2}\log\frac{\det K_K}{\det D_K} = -\frac{1}{2}\sum_{i=1}^{p_1+p_2}\log(1+\rho_i) \tag{3.61}$$

As seen in Eq.(3.61), KGV uses all eigenvalues of the decomposition, whereas KCCA uses only the maximum value. (Bach and Jordan, 2002) shows that the KGV is good at detecting independency. On the other hand, it estimates the mutual information with a bias for high dependence cases. In Figure 3.13, KGV method using a Gaussian kernel is applied to two variables that have known mutual information indicated by the dashed line. The data is rotated by different angles $\theta\epsilon[0, 90]$ and KGV is recalculated using different σ parameters of the kernel (dotted lines), chosen as $0.25, 0.5, 1, 2, 4$ and the solid line represents the KGV value where σ tends to zero. It can be seen that, KGV could detect the independence at $\theta = 30$, but as the true mutual information value increase, the KGV follows it with a bias.

Kernel ICA-KGV was claimed to be more successful compared to another widely used method Fast-ICA (Hyvarinen, 1999). This suggests that it can be also a good association estimator for EEG signals.

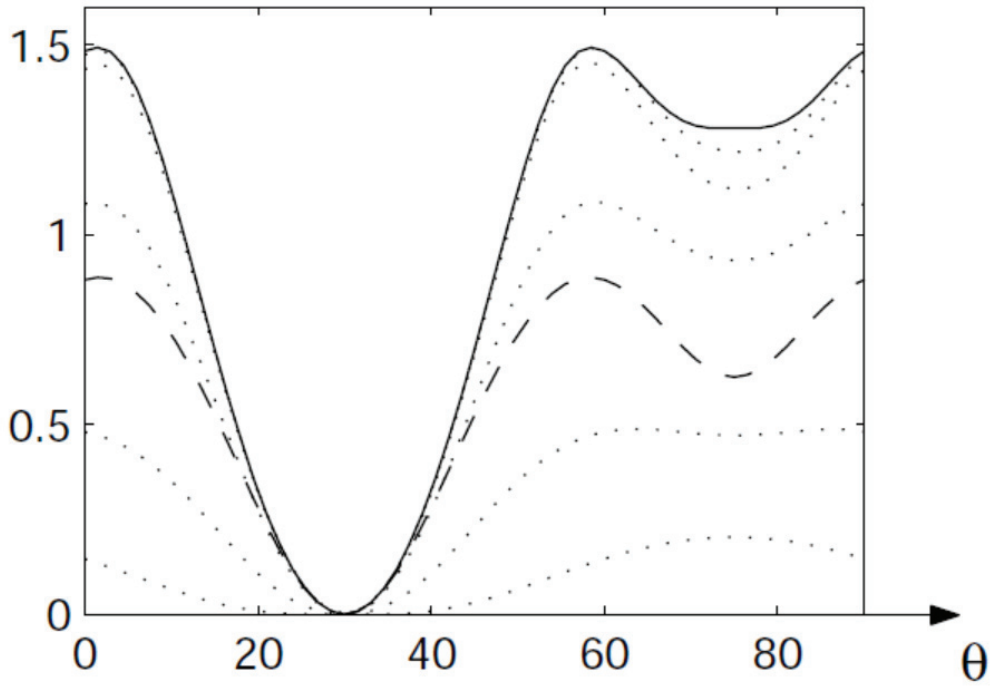


Figure 3.13. Kernel Generalized Variance values relation with mutual information for different kernel σ parameters and different θ (Source: Bach and Jordan (2002))

3.2.3.5. Rank-Based Renyi Information Estimation (REGO)

REGO estimates Renyi information using the relation between the Renyi entropy of copula function and the optimized Euclidean graphs for the variables defined in the interval [0,1] (Póczos et al., 2010). The method is completed in two steps:

- Find the empirical copula of variables
- Implement graph optimization and determine the minimum total graph length

Let X and Y be two random variables, as in Schweizer-Wolff method, and the variables U and V be defined as the cumulative distribution functions of X and Y, $U = F_X(X)$ and $V = F_Y(Y)$, while

$$C(u, v) = F(F_x^{-1}(u), F_y^{-1}(v)) \tag{3.62}$$

is copula function. It can be shown that the Renyi information of variables U and V is equal to the original variables X and Y (Vajda, 1989).

$$I_\alpha(X,Y) = I_\alpha(U,V) \tag{3.63}$$

To support this proposition, the relation between the copula and the mutual information is given in the appendix B.5. Since the new variables, U and V are standard uniform random variables, the marginal entropies are zero. Thus the joint entropy and the mutual information have the same magnitude values but reverse sign with $I_\alpha(U,V) = -H_\alpha(U,V)$, where Renyi α entropy is given by

$$H_\alpha(U,V) = \frac{1}{1-\alpha}\log\int_X f^\alpha(u,v)dudv \tag{3.64}$$

For the second step, the relation between the entropy and the total edge length L_n is required. Let G_n be a complete graph consisting of n nodes (for n sampled data) and $\mathcal{G}$ be a graph system of G_n. For the data $\mathbf{Z} = \{U,V\}$, the p power of the minimum total edge length is defined as

$$L_n(\mathbf{Z}) = \min_{G\epsilon\mathcal{G}} \sum_{((i,j)\epsilon E(G))} \left\|\mathbf{Z}^i - \mathbf{Z}^j\right\|^p \tag{3.65}$$

and the entropy estimator for n sampled dataset is given by

$$\hat{H}_n(\mathbf{Z}) = \frac{1}{1-\alpha}\log\frac{L_n(\mathbf{Z})}{\gamma_{d,\alpha}n^a} \tag{3.66}$$

In the expression above, $\gamma_{d,\alpha}n^a$ is defined as a universal constant that depends on the dimension d and α. (Redmond and Yukich, 1996) proved that the estimator converges to true entropy as the sample number $n \to \infty$. Different graph systems can be used for this method such as spanning tree graphs $(\mathcal{G}_{ST})$ (Hero and Michel, 1998), Hamiltonian cycles $(\mathcal{G}_H)$ and k-nearest neighbor graphs $(\mathcal{G}_{R(k)})$ (Póczos et al., 2010). A software that runs REGO algorithm by kNN graphs is published in http://www.cs.cmu.edu/ bapoczos/codes/REGO_with_kNN.zip. γ is empirically decided by implementing Monte-Carlo in the software. Besides (Pál et al., 2010) proved the $\mathcal{G} = \mathcal{G}_{R(k)}$ case.

3.2.4. Generalized Synchronization Index

In a system, with a driving variable x and a response variable y, a transformation relationship can be represented as $y(t) = \phi(x(t))$. For a simple transformation, the variables may tend to move on a line on the graphic. But for richer transformations or in a chaotic system, the variables are observed more on an orbit and on a geometrical shape.

The relationship on this trajectory is called *generalized synchronization* (Rulkov et al., 1995).

General synchronization is hard to detect by linear correlation methods. In order to address the issue (Arnhold et al., 1999) proposed to use the information of how the neighborhood of one effects the the neighborhood of the other variable to measure the mapping quality of transformation, named as *generalized synchronization index* or *nonlinear interdependence*. To measure generalized synchronization index of time series $x[n]$ and $y[n]$, two vectors are formed using the samples of x_n as $\mathbf{x_n} = \{x_n, ..., x_{n-(m-1)*\tau}\}$ and y_n as $\mathbf{y_n} = \{y_n, ..., y_{n-(m-1)*\tau}\}$. In this notation, m is the embedding dimension and τ is time lag value. k-nearest neighbor indices of the vectors $\mathbf{x_n}$ and $\mathbf{y_n}$ are represented with $r_{n,j}$ and $s_{n,j}$ respectively, where $j = 1, 2, ..., k$. The mean squared euclidean distance of the vector $\mathbf{x_n}$ is then

$$R_n^{(k)}(X) = \frac{1}{k}\sum_{j=1}^{k}\left(\mathbf{x_n} - \mathbf{x_{r_{n,j}}}\right)^2 \tag{3.67}$$

Y conditioned mean square distance of (x_n) can be calculated using the kNN indices of (y_n) as

$$R_n^{(k)}(X|Y) = \frac{1}{k}\sum_{j=1}^{k}\left(\mathbf{x_n} - \mathbf{x_{s_{n,j}}}\right)^2 \tag{3.68}$$

Using the indices of the other variable always causes bigger distance $R_n^{(k)}(X|Y) \geq R_n^{(k)}(X)$. But the expected value of the two calculations are close to each other with respect to dependency. Thus a measure of dependence is defined by the average rate of $R_n^{(k)}(X)$ and $R_n^{(k)}(X|Y)$ as

$$S^{(k)}(X|Y) = \frac{1}{N}\sum_{j=1}^{N}\frac{R_n^{(k)}(X)}{R_n^{(k)}(X|Y)} \tag{3.69}$$

The expression in Eq.(3.69) is called the *S index*. The range of S is between 0 and 1. The index gets a value of 1 in full dependence, and 0 in independence cases. In effect, $S^{(k)}(X|Y)$ estimates the mapping quality of X on the variable Y. However, the inverse values ($R_n^{(k)}(Y)$, $R_n^{(k)}(Y|X)$ and $S(k)(Y|X)$) may not necessarily give the same value. The value which should be accepted depends on which is the driving variable and which is the response variable.

Other indices related to generalized synchronization method are shown in Table 3.2. Following the instruction of the S index, the H index is proposed (Arnhold et al.,

Table 3.2. Generalized Synchronization Indices

Reference	Equation	Remarks
(Arnhold et al., 1999)	$S^{(k)}(X\|Y) = \frac{1}{N}\sum_{n=1}^{N} \frac{R_n^{(k)}(X)}{R_n^{(k)}(X\|Y)}$	Normalized
(Arnhold et al., 1999)	$H^{(k)}(X\|Y) = \frac{1}{N}\sum_{n=1}^{N} \log \frac{R_n(X)}{R_n^{(k)}(X\|Y)}$	Not normalized May be negative Robust against noise
(Quiroga et al., 2002)	$N^{(k)}(X\|Y) = \frac{1}{N}\sum_{n=1}^{N} \frac{R_n(X)-R_n^{(k)}(X)}{R_n(X)}$	Normalized Robust against noise May be negative
(Andrzejak et al., 2003)	$M^{(k)}(X\|Y) = \frac{1}{N}\sum_{n=1}^{N} \frac{R_n(X)-R_n^{(k)}(X)}{R_n(X)-R_n^{(k)}(X\|Y)}$	Normalized
(Chicharro and Andrzejak, 2009)	$L^{(k)}(X\|Y) = \frac{1}{N}\sum_{n=1}^{N} \frac{G_n(X)-G_n^{(k)}(X)}{G_n(X)-G_n^{(k)}(X\|Y)}$	Normalized Rank Based

1999). H index calculates the geometrical average rather than arithmetical one by calculating the average logarithms of $R_n^{(k)}(Y)/R_n^{(k)}(X|Y)$, and offers robustness to outliers. On the other hand, proper normalization is lost. $R_n(X)$ is the mean in distance of $k = N-1$ case as

$$R_n(X) = \frac{1}{N-1}\sum_{j=1}^{N-1} \left(\mathbf{x_n} - \mathbf{x_{r_{n,j}}}\right)^2 \tag{3.70}$$

The other two methods (Table 3.2), N and M indices are more robust against noise than S index again and they are also normalized. Finally, L index is the ranked based version of S index and $G_n^{(k)}(X)$ represents the rank distances between $\mathbf{x_n}$ and its k neighbors.

CHAPTER 4

EXPERIMENTAL SETUP AND RESULTS

Dependence measures were compared in three different aspects. First, methods were compared in terms of their computational time. A second analysis was carried out to evaluate their accuracy-time performance. Finally, methods were analyzed as measures of dependence for EEG signals.

Various studies evaluate the performance of connectivity measures in different ways. (Lee and Hsieh, 2014) compared three connectivity indices by emotional states classification accuracy. (Silfverhuth et al., 2012) implemented a comparison of six different connectivity measure methods, by evaluating the ability to differentiate the direct causal and non-causal connections on simulated EEG data which has the predefined connections between the electrodes. The connection types were assessed by Chi-square tests. Similarly (Astolfi et al., 2007) compared three cortical connectivity estimators by evaluating reconstruction abilities for a generated EEG signal which have known connection strengths. (Haufe and Ewald, 2016) used three different dependence measures on generated EEG signal again. On the other hand, as in (Fallani et al., 2014), graph analysis is the most widely used way to analyze connectivity measure methods.

In our case, we consider the performance of these measures based on the statistical significance of the difference between connectivity structures of right-left fist, both fist-both feet, right-left fist imagery and both fist-both feet imagery tasks. From this point of view, firstly a hierarchical clustering based procedure was proposed for the analysis of measure methods. Three different signal epochs at [0.2-0.5], [0.2-1] and [0.2-2] sec intervals were used for the analysis. Secondly, relations of channels placed in positions responsible for motor tasks were analyzed in the scope of statistically significance again. Finally, experiments were repeated for 5 selected methods on [0.2-4.1] sec time interval.

4.1. Computation Time Comparison

There is one thing obvious for all dependence methods in this study: the computation time of any method is unrelated to data content but the data size. So the computational times tests are focused on comparing the methods according to the response time under

variable data size. Test were performed on MATLAB © (1 Apple Hill Drive Natick, MA 01760-2098 USA) using 16 channel dummy data generated randomly with variable sample number: 250, 300, 350, ..., 750. For every sample size each method was run 100 times on the same input data. Slowest 25 time values were removed to avoid possible slowdown caused by other processes run in computer.

A white paper including recommendations to perform a healthier test on MATLAB was published earlier (McKeeman, 2016). A series of steps in that paper were followed accordingly to run the test in a stable environment:

- Kill every process possible. Turn off any anti-virus programs.
- Close the internet and other local area connections.
- Unplug or disconnect external devices (mass storage devices and such)
- Allocate just one core for Matlab (by set affinity option in Windows). While this helps obtain a stable test, it also disables parallel working feature of some Matlab functions.
- Disable the boosting feature of CPU.
- Disable the automatic sleep or turn off functions of the screen.
- Do not read or write files, do not access the input data or print output.

MATLAB configuration and computer features are given in Table 4.1. In addition, appendix C.1 shows the detailed procedure applied on Powershell application.

In Figure 4.1, time performance of methods are visualized in three groups related to the former chapter. Generalized synchronization method is placed along with information theoretic methods. Correlation based methods (Spearman ρ, Kendall τ, SCC) and phase synchronization based methods (PLV, PLI) take less time as expected because of their simplicity. Furthermore, both methods proposed in this study, UV parametrization and data fitting based estimation methods compared favorably against the alternatives. However kNN based methods like Kraskov, REGO and the copula based method of Schweiser-Wolff showed a quadratically increasing time characteristic with increasing number of samples. Finally, although the generalized synchronization index method performs well in low sample sizes, it consumes more and more time in search of neighbors even in reasonable embedding dimensions (in this study m is chosen as 10) as the sample size increases.

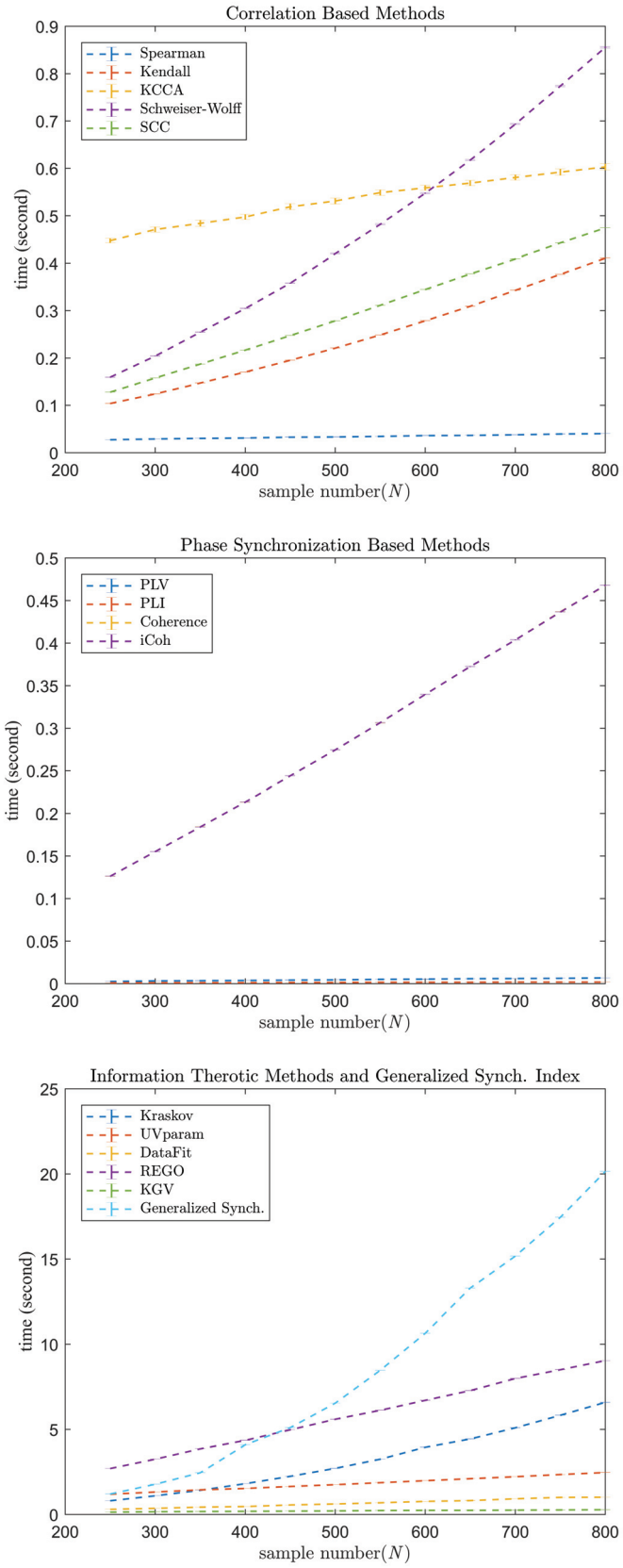


Figure 4.1. Computation times of measures with varying sample size

Table 4.1. MATLAB configuration and PC features

MATLAB version	9.2.0.538062 (R2017a)
MATLAB root	C:\Program Files\MATLAB\R2017a
MATLAB accelerator	enabled
MATLAB JIT	enabled
MATLAB assertions	disabled
MATLAB Desktop	enabled
Java JVM	enabled
Java version	Java 1.7.0 60-b19 with Oracle Corporation Java HotSpot(TM) 64-Bit Server VM mixed mode
CPU	x86 Family 6 Model 158 Stepping 9, GenuineIntel
Operating Sys.	Microsoft Windows 10 Enterprise
Number of cores	4
Number of threads	4

4.2. Time-Accuracy Performance Analysis of Information Theoretic Methods

In order to compare the information theoretic methods, different bivariate variables were generated by applying a linear transformation on the independent pair of variables with Gaussian-Gaussian, Gaussian-uniform and Gaussian-exponential distributions. Rotation vectors with $\{\{0,0\},\{10,-10\},\{20,-20\},\{30,-30\},\{40,-40\}\}$ angle pairs were used as transformation vectors. In this way, bivariate variables with known mutual information values and different degrees of dependence were obtained.

The experiments were repeated 100 times for each configuration of angle and sample size, varying from 250 to 750. All information theoretic methods (Kraskov, data fitting, UV parametrization, REGO and KGV) were applied on this set of data and the results were compared. During this test, k was chosen as 1 for all kNN methods (Kraskov and REGO). Additionally, α parameter was defined as 0.99 since mutual information was computed and γ parameter of REGO was determined using Monte Carlo method to match the α and k values (Póczos et al., 2010). Running mean smoothing method was used for variance estimation in data fitting method, and finally, $\sigma = 5$ was selected empirically for KGV.

Firstly, the methods were compared for varying dependence degrees (Figure 4.2). The average mutual information values and the corresponding standard deviations for each angle are shown in Figure 4.2 for methods. The sample size N was chosen as 350 for Figure 4.2 in accordance with the subsequent experiments. Finally, the mean squared

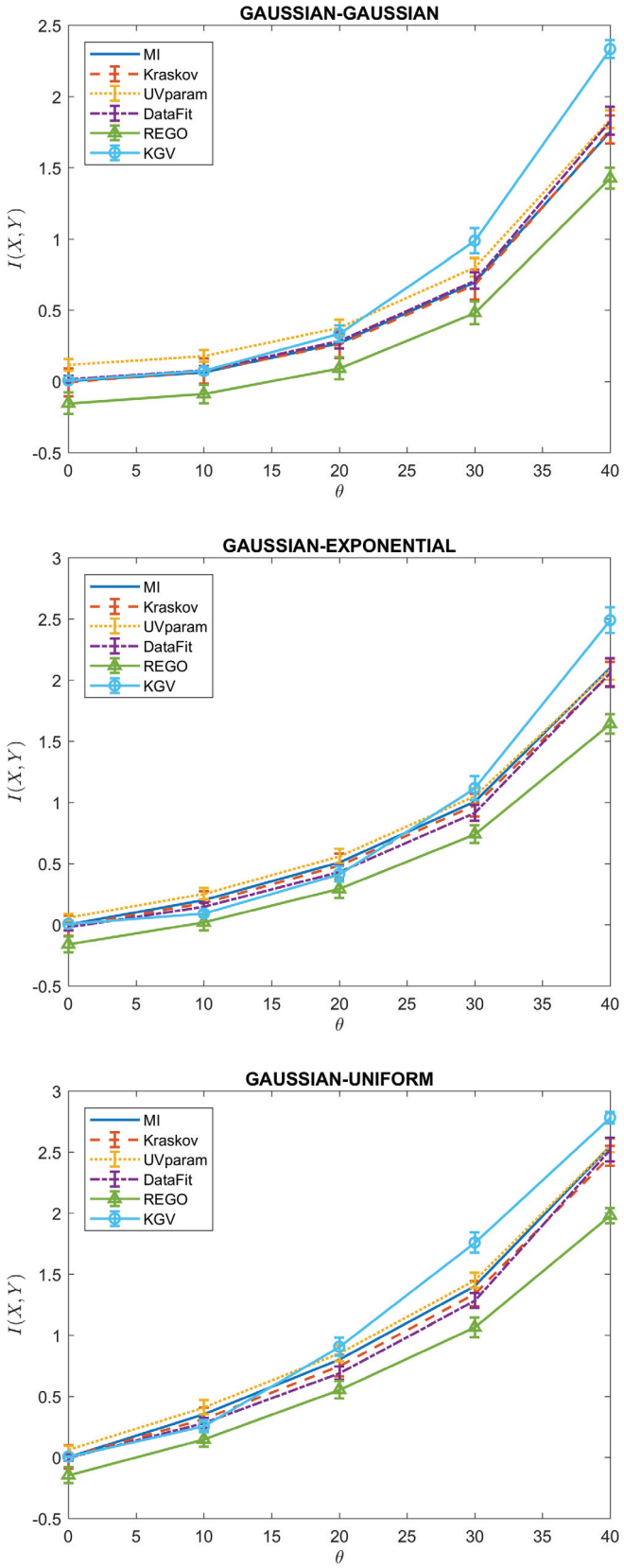


Figure 4.2. Mutual information values of measures with varying distributions and dependency

error values of all methods in all conditions (varying sample number, angle and density) against the average time performances are give in Figure 4.3.

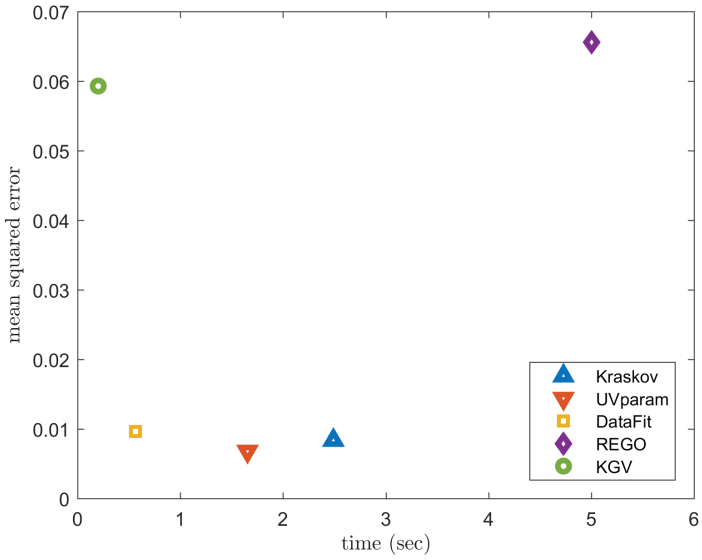


Figure 4.3. Time-Accuracy comparison of methods

4.3. Hiearchical Clustering Based Evaluation

As discussed earlier, brain connectivity models are constructed and evaluated typically using graph-based methods. However, a thresholding process on the connectivity matrice is a necessity for these methods and the threshold parameter affects the resulting connectivity structure and significance. Since the association values are not normalized for all connectivity measure methods, defining the optimum threshold parameter for each method is a difficult problem.

In this study, a hierarchical clustering (HiCl) based connectivity group selection method are proposed as in (Mammone et al., 2018). In each clustering step, obtained clusters are evaluated and most statistically significant one is chosen among candidate connectivity clusters. The comparison in this experiment is based on the ability of statistically discriminating the four event pairs given below:

- left fist-right fist
- left fist imagery-right fist imagery

- both fist - both feet
- both fist imagery - both feet imagery

Let $\mathbf{S}_1$ and $\mathbf{S}_2$ be the collections of epochs for such as left fist and right fist, respectively. Then the procedure for each measure method on these collections are applied as bellow:

1. Compute connectivity matrix for each epoch in $\mathbf{S}_1$ and $\mathbf{S}_2$:

   $\mathbf{A}_k^1$: $n \times n$ connectivity matrix of the k^{th} epoch in $\mathbf{S}_1$ for n channel EEG data

   $\mathbf{A}_k^2$: $n \times n$ connectivity matrix of the k^{th} epoch in $\mathbf{S}_2$ for n channel EEG data

2. Calculate the average values for all channel pairs in each event type. Let $a_{i,j,k}^r$ is the connectivity measure value between i^{th} and j^{th} channels for the k^{th} epoch of the r^{th} event. Then,

$$\langle a_{i,j} \rangle_1 = \frac{1}{|I_1|} \sum_{k \epsilon I_1} a_{i,j,k}$$
$$\langle a_{i,j} \rangle_2 = \frac{1}{|I_2|} \sum_{k \epsilon I_2} a_{i,j,k}$$

   where $\mathbf{I}_1$ and $\mathbf{I}_2$ are the epoch index group for each event.

3. Apply hierarchical clustering (HiCl) on average connectivity matrices $\langle \boldsymbol{A} \rangle_{\mathbf{1}}$ and $\langle \boldsymbol{A} \rangle_{\mathbf{2}}$, defined as

$$(\langle A \rangle_1)_{i,j} = \langle a_{i,j} \rangle_1$$

   and

$$(\langle A \rangle_2)_{i,j} = \langle a_{i,j} \rangle_2$$

   Collect the clusters obtained in the successive steps of the clustering process and group them as $\mathbf{J}_1$ and $\mathbf{J}_2$ (Figure 4.4a).

4. For each cluster in $\mathbf{J}_1$, find the sub-matrix that consist of this cluster's elements in each connectivity matrice $\mathbf{A}_k^1$ and $\mathbf{A}_k^2$ which belongs to k^{th} epoch. Get the minimum value of these sub-matrices and collect them in two vectors.

5. Apply one-tailed student's t-test between these two vectors and get the p values (Figure 4.4b).

$$p_1^l = t\text{-test} \begin{cases} \{min\left\{a_{i,j,k}; i, j \epsilon J_l^1\right\}, k \epsilon I_1\} \\ vs \\ \{min\left\{a_{i,j,k}; i, j \epsilon J_l^1\right\}, k \epsilon I_2\} \end{cases}$$

6. Repeat step 4-5 for $\mathbf{J}^2$.

$$p_2^l = t\text{-test} \begin{cases} \{min\left\{a_{i,j,k}; i, j \epsilon J_l^2\right\}, k \epsilon I_1\} \\ vs \\ \{min\left\{a_{i,j,k}; i, j \epsilon J_l^2\right\}, k \epsilon I_2\} \end{cases}$$

7. Find the minimum p value in $\mathbf{p}_1$ and $\mathbf{p}_2$ and the cluster that provides this minimal p value to represent the connectivity measure in question.

As seen in the procedure, one-tailed student's t-test is used to evaluate the statistical significance. For the analysis of clusters in $\mathbf{J}_1$, the null hypothesis H_0 is that the mean value of the vector obtained from $\mathbf{A}_k^1$ is less than or equal to the mean value of the vector obtained from $\mathbf{A}_k^2$. The alternative hypothesis H_a is that the mean value of the the vector obtained from $\mathbf{A}_k^1$ is greater than the mean value obtained from $\mathbf{A}_k^2$. Since the clusters in $\mathbf{J}_1$ are created by applying HiCl on $\langle \boldsymbol{A} \rangle_1$, it is expected that the connectivity measures should be greater for the first event. As for clusters in $\mathbf{J}_2$, the null hypothesis and the alternative hypothesis are reversed.

While applying HiCl method on the connectivity matrix, averaging, choosing minimum, maximum and some other parameters can be used to compute distance between clusters. Since the purpose in brain connectivity studies is to detect dense cluster relations in all pairwise electrodes rather than strong individual relations, minimum association values are chosen to represent the cluster distance values. Besides, for student's t-test, the minimum values are chosen to represent connectivity sub-matrix of each cluster and epoch as explained in the procedure.

Hierarchical clustering based procedure was applied to 103 subjects after removing 6 damaged and inappropriate subject data. Recorded signals at [0.2-0.7], [0.2-1.2] and [0.2-2.2] sec time intervals were used for comparison tests. Average p-values of 103 subjects for [0.2-1.2] time interval are given in Table 4.2. Geometrical averaging method was preferred rather than arithmetical averaging, because larger p-values tend to mask the other values in arithmetical averaging method.

A p-value of 0.05 for a method in evaluation of a task means that there is a %5 probability of facing the observed data, when the null hypothesis is true. This is a low probability and it would be reasonable to reject that null hypothesis. But after repeating

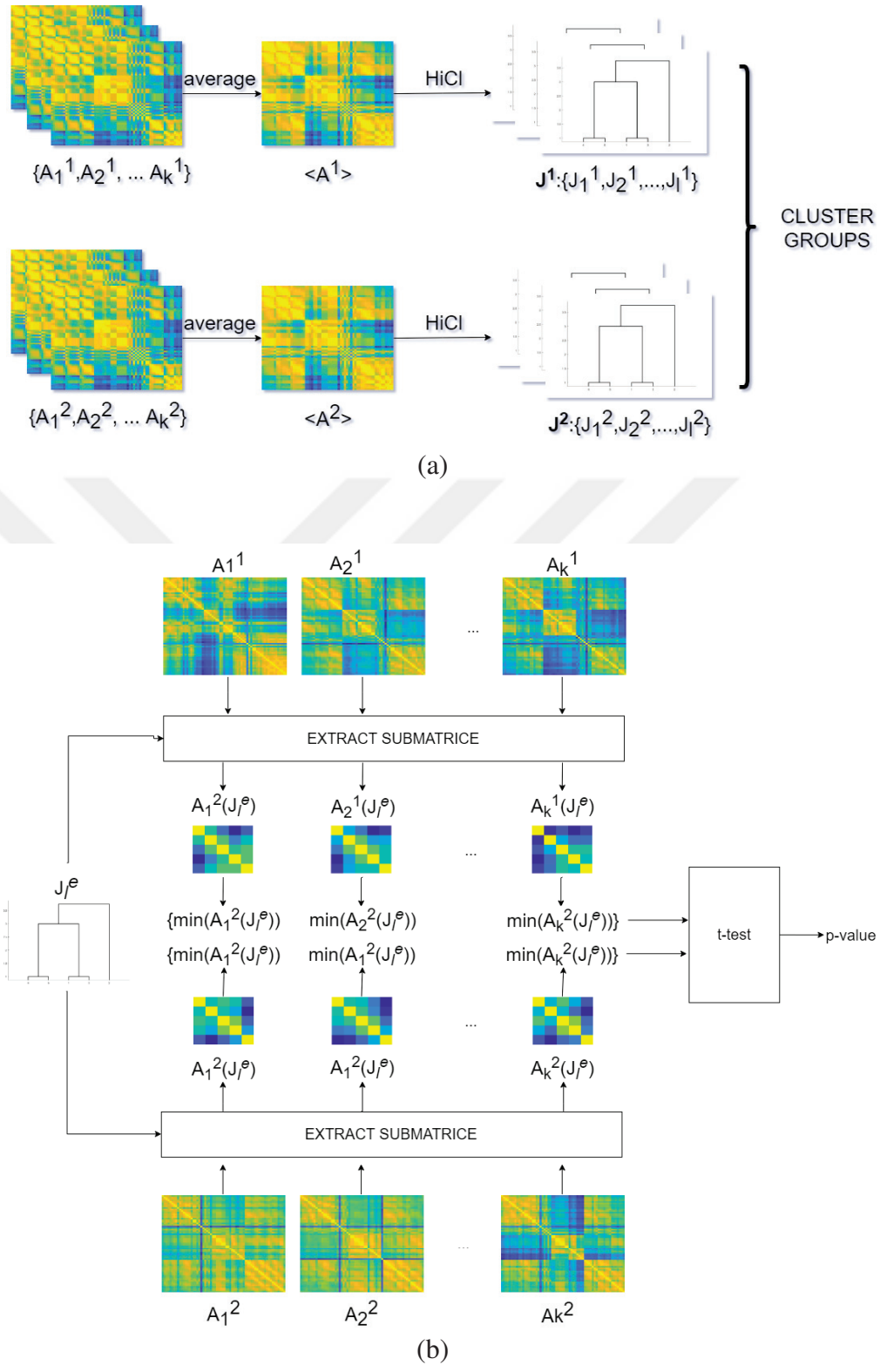


Figure 4.4. Hierarchical clustering process a) process of determining the candidate clusters b) calculation of p-value for a cluster

Table 4.2. Average p-values across subjects without multiple comparison correction

	Left Fist Right Fist	Left Fist Imag. Right Fist Imag.	Both Feet Both Fist	Both Feet Imag. Both Fist Imag.
Spearman	0.0013	0.0016	0.0003	0.0025
Kendall	0.0012	0.0016	0.0002	0.0021
Kraskov	0.0017	0.0025	0.0005	0.0029
Uvparam	0.0013	0.002	0.0003	0.0028
DataFit	0.0016	0.0019	0.0003	0.0028
REGO	0.0019	0.0023	0.0004	0.0027
KGV	0.0029	0.0025	0.0005	0.0034
KCCA	0.0017	0.0017	0.0003	0.0027
SW	0.0013	0.0017	0.0003	0.0024
PLI	0.0007	0.0009	0.0005	0.0011
PLV	0.0013	0.0011	0.0003	0.0021
SCC	0.0024	0.0021	0.0009	0.002
Coh	0.0023	0.0019	0.0007	0.0025
iCoh	0.0009	0.001	0.0012	0.0014
Gen. Sync	0.0083	0.0055	0.0013	0.0081

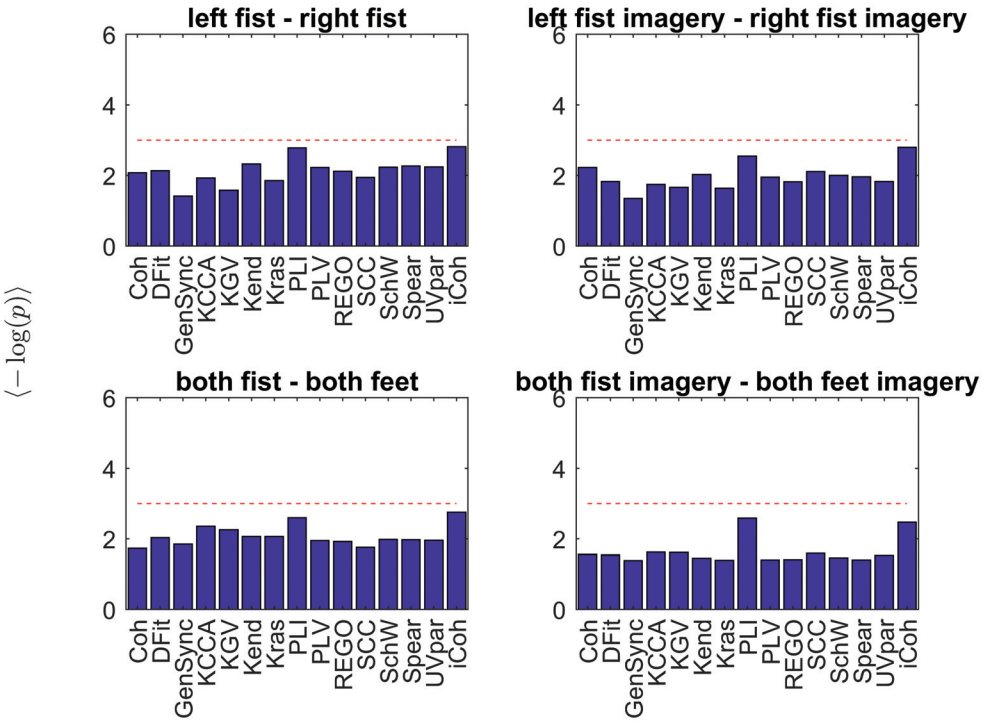
the test for 124 clusters as in the procedure, it is more likely to observe improbable results even if the method is not significant due to a multiple comparison problem. Furthermore, choosing the most significant cluster is like rolling a die 100 times and expecting one of them to be 6, except that the tests in this procedure are not independent. Because of that, a correction method is needed so that the p-values can be adjusted properly.

The most widely used and the simplest way of p-value adjustment is Bonferroni correction. The method finds corrected p-values by dividing each p-value by the number of experiment as $\tilde{p}(i) = kp_i$. In relation to Bonferroni, Sidak calculates the corrected p-values as $\tilde{p}(i) = 1 - (1 - p_i)^k$ (Westfall and Young, 1993). Since both methods are excessive especially for non-independent multiple experiments as in this case, Benjamini-Hochberg method was used to adjust p-values (Benjamini and Hochberg, 1995). Benjamini-Hochberg method is implemented in to steps.

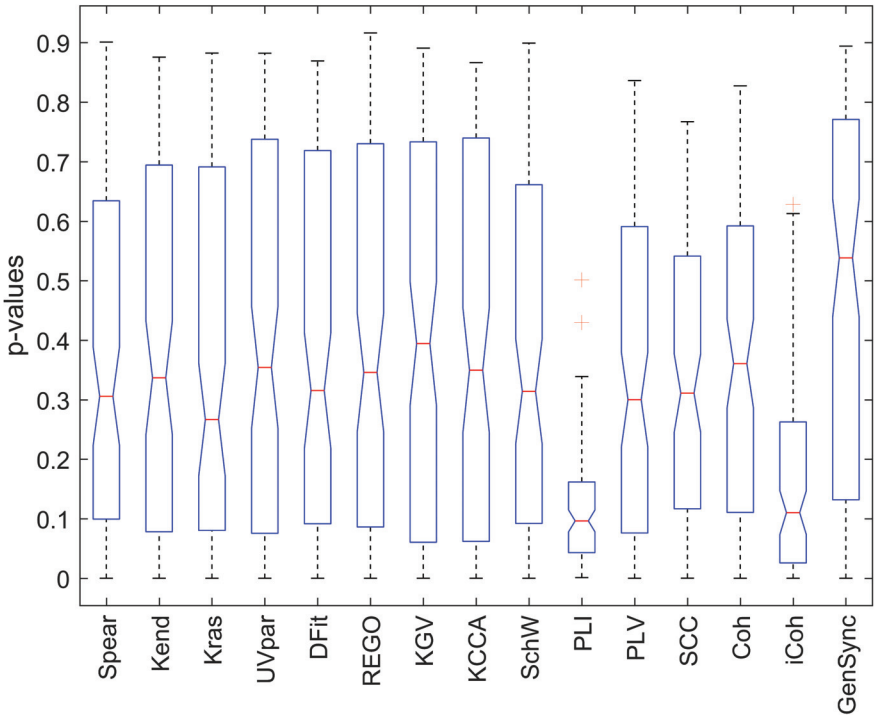
1. Sort each p-value in ascending order. Each sorted p-value $p(1)$,$p(2)$,...,$p(k)$ has a rank. The rank of smallest p-value $p(1)$ is 1, rank of second one $p(2)$ is 2 and the largest p-value $p(k)$ has the rank of k.

2. Multiply each p-value by k/i and find the minimum p-value.

$$\tilde{p} = \min_i p(i)\frac{k}{i} \tag{4.1}$$

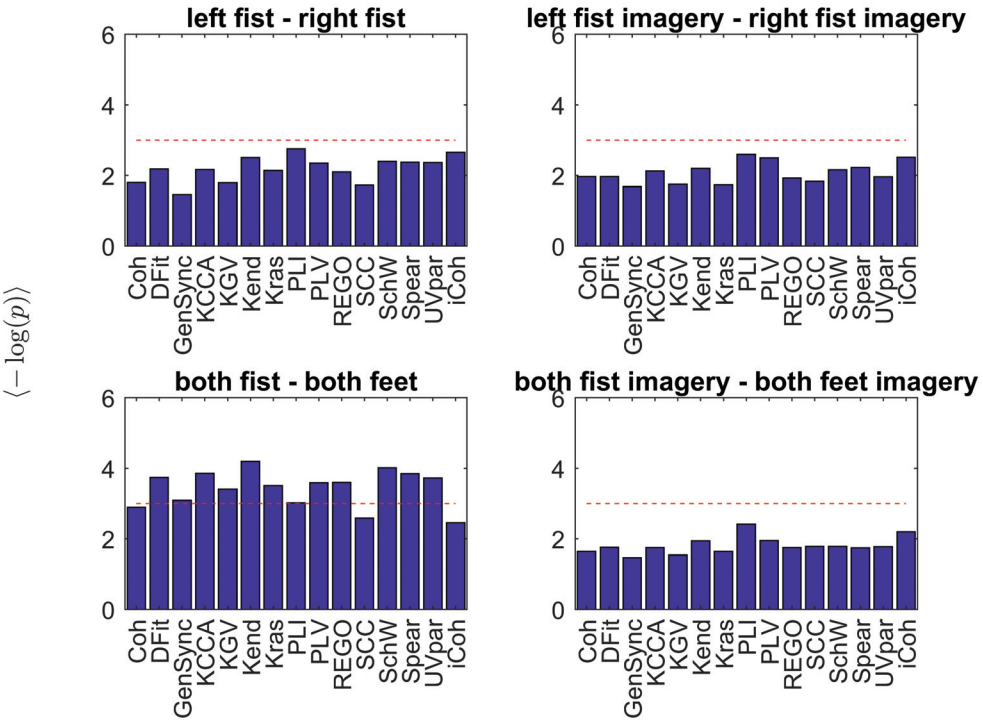


(a)

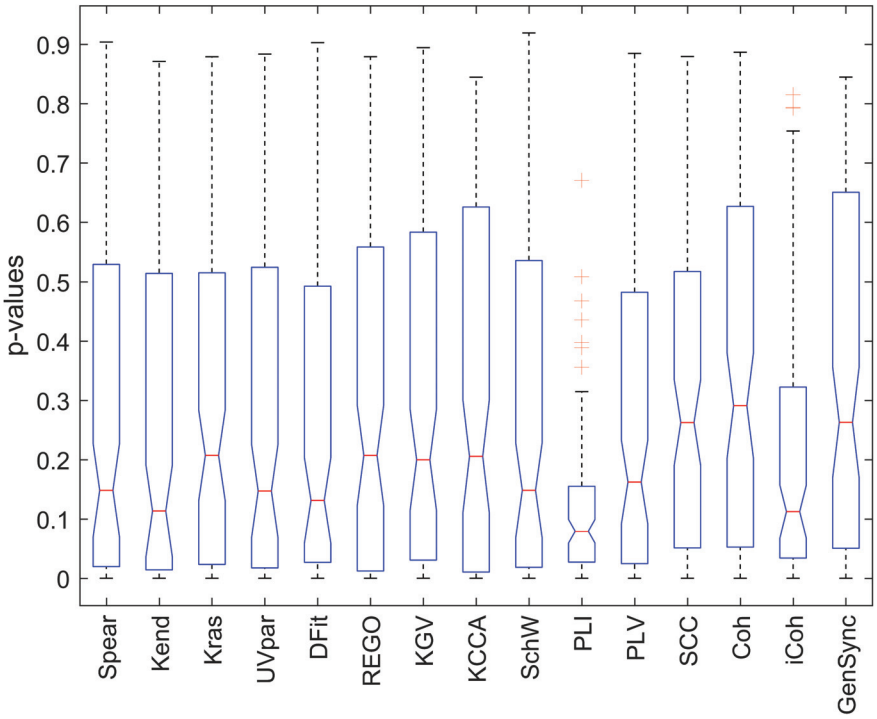


(b)

Figure 4.5. For time interval [0.2-0.7] sec a) average of logarithm of p-values for each task b) p-values box plot of *both fist - both feet* task

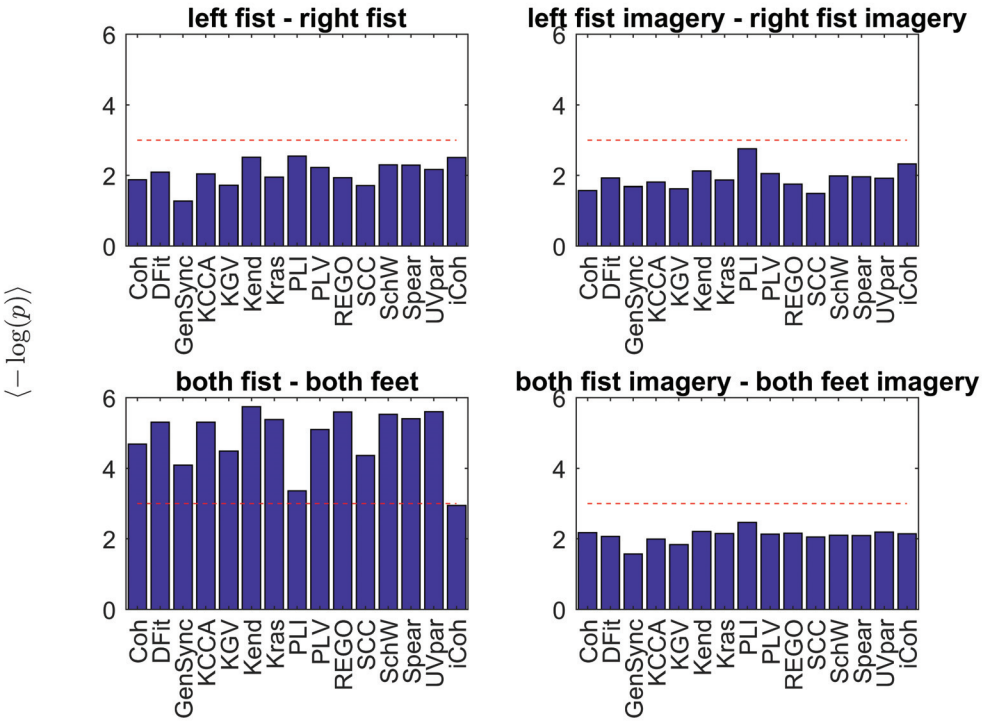


(a)

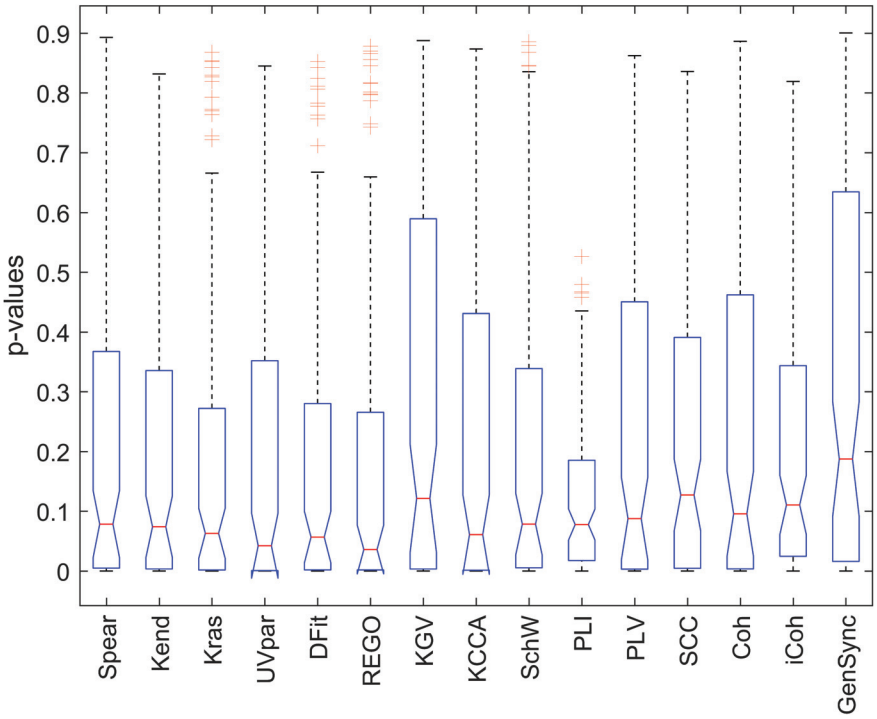


(b)

Figure 4.6. For time interval [0.2-1.2] sec a) average of logarithm of p-values for each task b) p-values box plot of *both fist - both feet* task



(a)



(b)

Figure 4.7. For time interval [0.2-2.2] sec a) average of logarithm of p-values for each task b) p-values box plot of *both fist - both feet* task

Table 4.3. Average p-values of subjects after multiple comparison correction

	Left Fist Right Fist	Left Fist Imag. Right Fist Imag.	Both Feet Both Fist	Both Feet Imag. Both Fist Imag.
Spearman	0.0928	0.1081	0.0212	0.1739
Kendall	0.0816	0.1106	0.0151	0.1426
Kraskov	0.1175	0.1763	0.03	0.1927
Uvparam	0.0941	0.1403	0.0242	0.1695
DataFit	0.1123	0.1399	0.0237	0.1711
REGO	0.1225	0.1452	0.0272	0.1728
KGV	0.1663	0.1726	0.0331	0.2131
KCCA	0.114	0.119	0.0212	0.1726
SW	0.0904	0.1156	0.018	0.168
PLI	0.0636	0.0745	0.0487	0.0893
PLV	0.0951	0.082	0.0275	0.1413
SCC	0.1773	0.1598	0.075	0.1667
Coh	0.1642	0.1391	0.0554	0.1927
iCoh	0.0699	0.0809	0.0854	0.1105
Gen. Sync	0.2331	0.1846	0.0453	0.2306

The geometric average of the corrected p-values is given in Table 4.3, following Benjamini-Hochberg adjustment. For a visual illustration, the arithmetical average of the logarithms of the corrected p-values are shown for each task (Figure 4.5a, Figure 4.6a, Figure 4.7a). In these figures, red dashed lines show the negative logarithm of 0.05 p-value. Note that significant differences in cluster connectivity are observed especially for *both fist-both feet* task. Box plots of this task provide more information about the density of values (Figure 4.5b, Figure 4.6b, Figure 4.7b). In box plot figures:

- Central marks for each method show the median p-values.
- Upper and lower borders of boxes indicate the 25^th and 75^th percentile of values.
- '+' markers indicate the outlier samples.
- Whiskers indicate the most extreme sample values that are not considered as outlier.

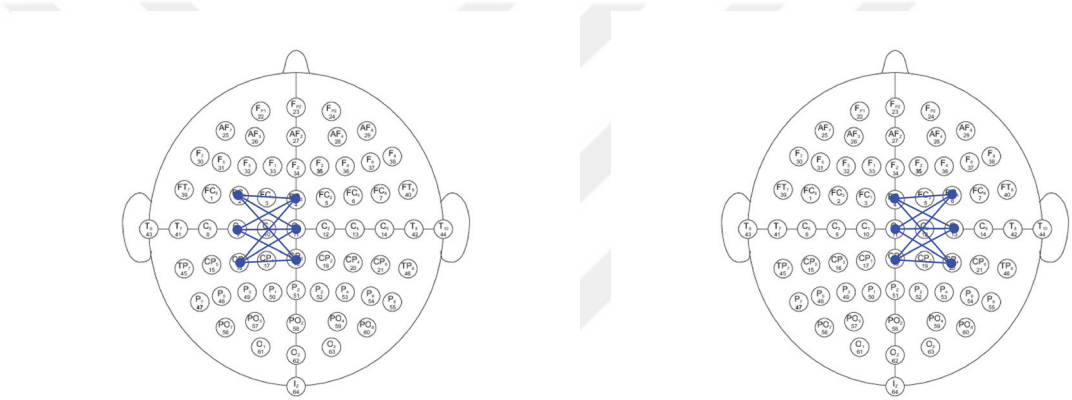
4.4. Evaluation on Selected Electrode Pairs

There are specific electrodes or electrode pairs used in BCI and conntectivity studies especially to discriminate motor and motor imagery tasks. (Nolte et al., 2004) evaluted the performance of iCOH method on C3-C4 electrode pair. In addition, another study that

also uses the Physionet data, implemented a BCI embedded system. Electrode pairs in the { FC5,Fz,C3,C4} cluster (Tummala et al., 2018).

In order to evaluate the significance of the coherence between these selected EEG channels, we carried out an additional significance test applied on the coherence measures obtained by the implemented methods. As used in (Loboda et al., 2014), the electrodes given below were chosen for this experiment:

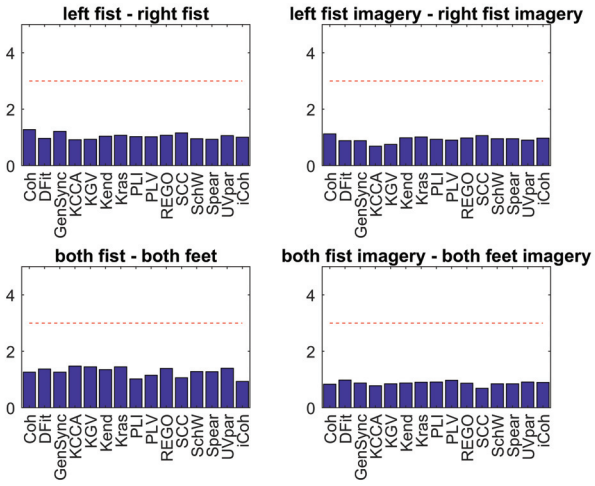
- For the midline area : FCz, Cz, CPz
- For the left hemisphere: FC3, C3, CP3
- For the right hemisphere: FC4, C4, CP4



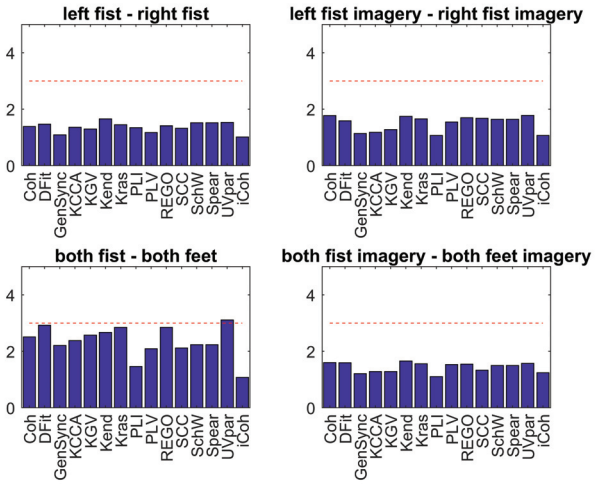
(a) (b)

Figure 4.8. Electrode pair mappings a) midline-left hemisphere b) midline-right hemisphere

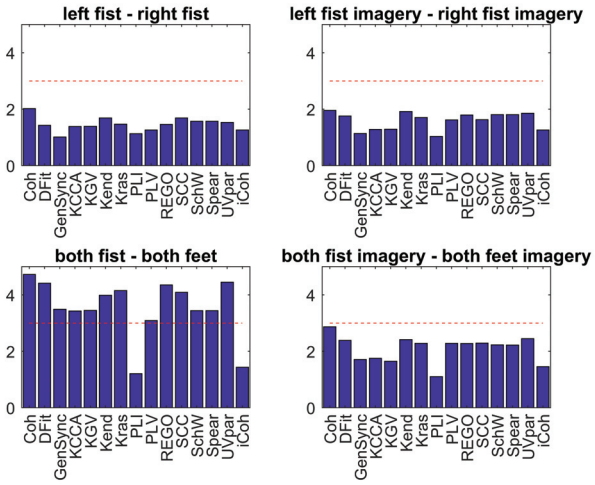
All possible electrode pairs were selected between *midline-left hemisphere* and *midline-right hemisphere* group pairs (Table 4.4). Then, significance tests were applied independently for these 18 pairs (Figure 4.8). Finally, the smallest p-value of the most significant pair was saved for each subject. similar to the hierarchical clustering experiment earlier, the Benjamini-Hochberg multiple comparison correction procedure was applied for the p-values of all 18 pairs before selecting the most significant one, and average of p-values logarithm were computed for signal durations at 0.5, 1 and 2 seconds (Figure 4.9).



(a)



(b)



(c)

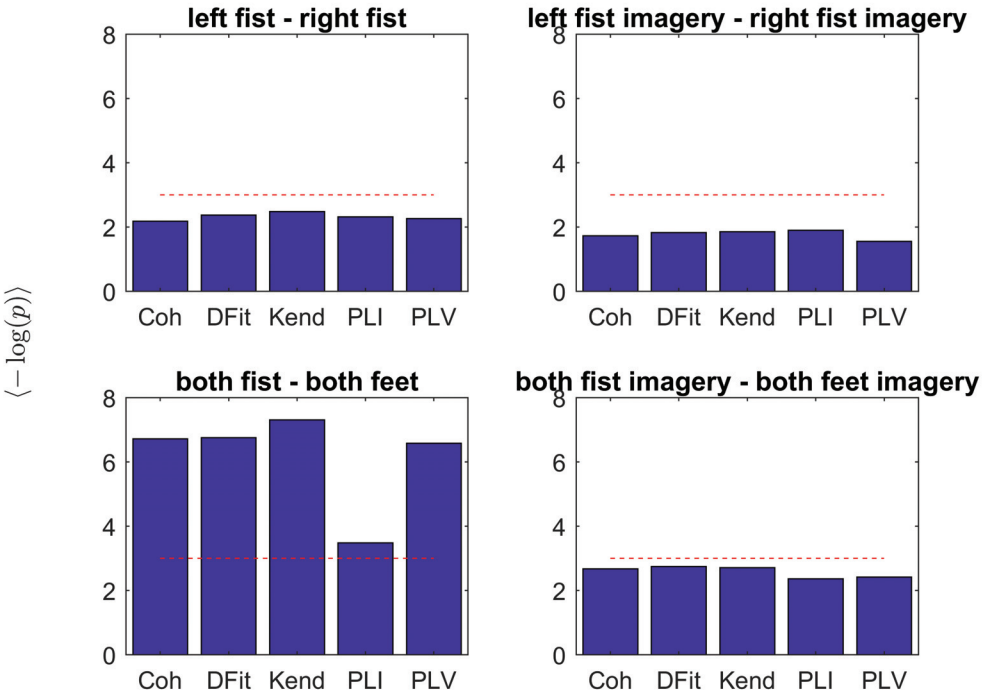
Figure 4.9. Average logarithmic p values for a) 0.5, b) 1 and c) 2 seconds duration

Table 4.4. Electrode pairs used for evaluation

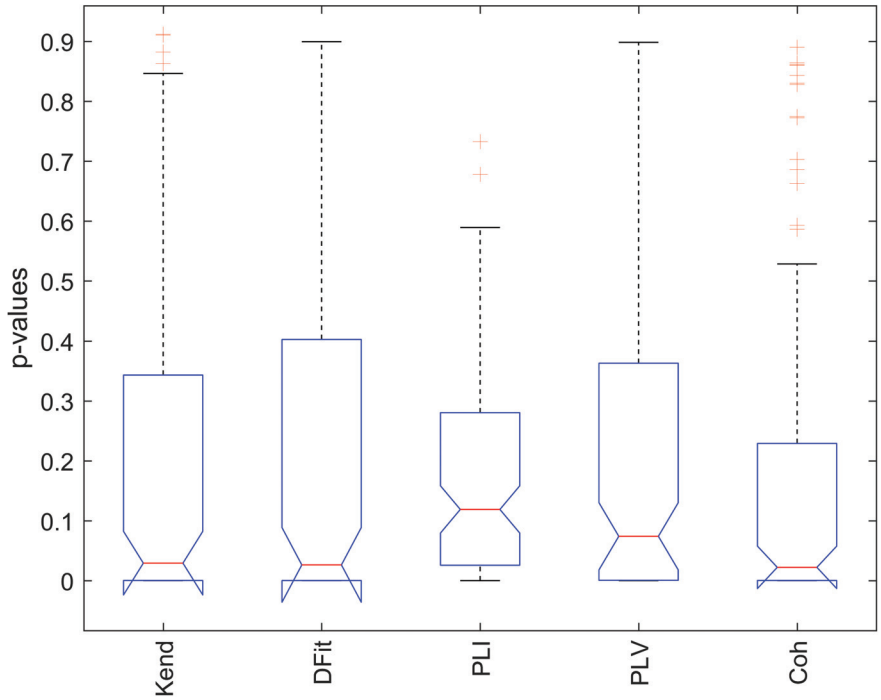
Pair Index	Midline	Left	Pair Index	Midline	Right
1	FCz	FC3	10	FCz	FC4
2	FCz	C3	11	FCz	C4
3	FCz	CP3	12	FCz	CP4
4	Cz	FC3	13	Cz	FC4
5	Cz	C3	14	Cz	C4
6	Cz	CP3	15	Cz	CP4
7	CPz	FC3	16	CPz	FC4
8	CPz	C3	17	CPz	C4
9	CPz	CP3	18	CPz	CP4

4.5. Analysis of Selected Methods at Longer Time Interval

Both experiments on EEG signals described above showed that, especially for *both fist-both feet* task, smaller p-values were observed with increasing observation time. Because of that, a final analysis was carried out on the full [0.2-4.1] second interval following the same procedures. For this experiment, 5 out of 15 methods were chosen based on their category, time performance and p-value performance characteristics. Data fitting based MI calculator were chosen to represent information theoretic methods because of its high p-values and low computational time. Kendall coefficient that had best p-values was chosen from the correlation based methods. PLI method was added to this experiment among the two methods that claim to avoid volume conduction effects, and finally, coherence and PLV methods were included in the experiment due to their increasing p-value characterics with increasing time interval. As mentioned above, hierarchical clustering evaluation and analysis on selected pairs were implemented for these methods on [0,2-4.1] sec interval. Logarithmic average and boxplot of subject p-values are shown in Figure 4.10 and Figure 4.11.

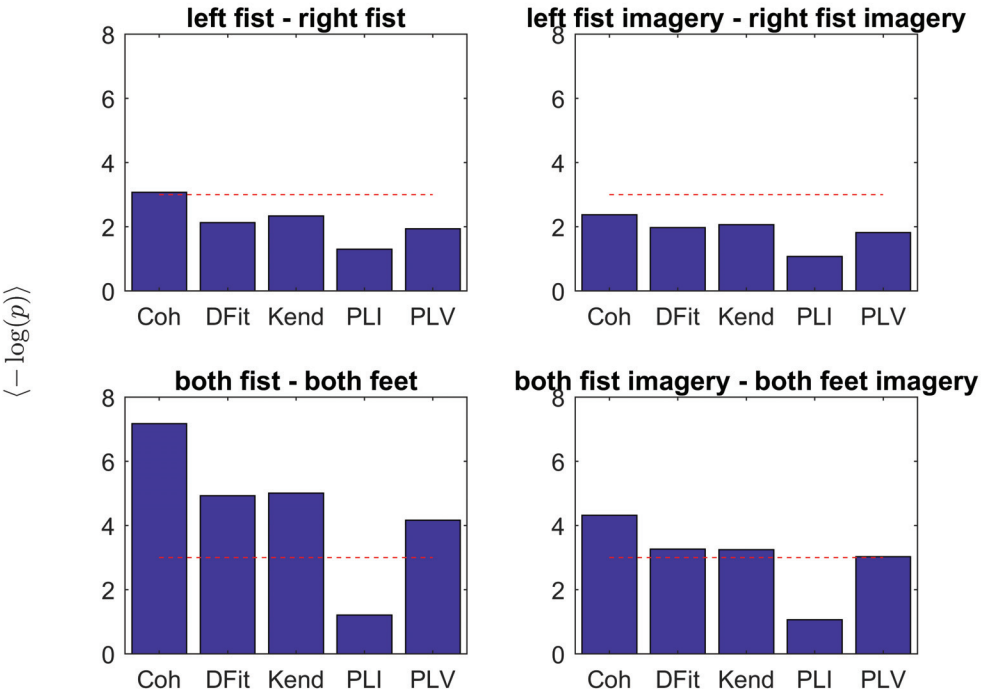


(a)

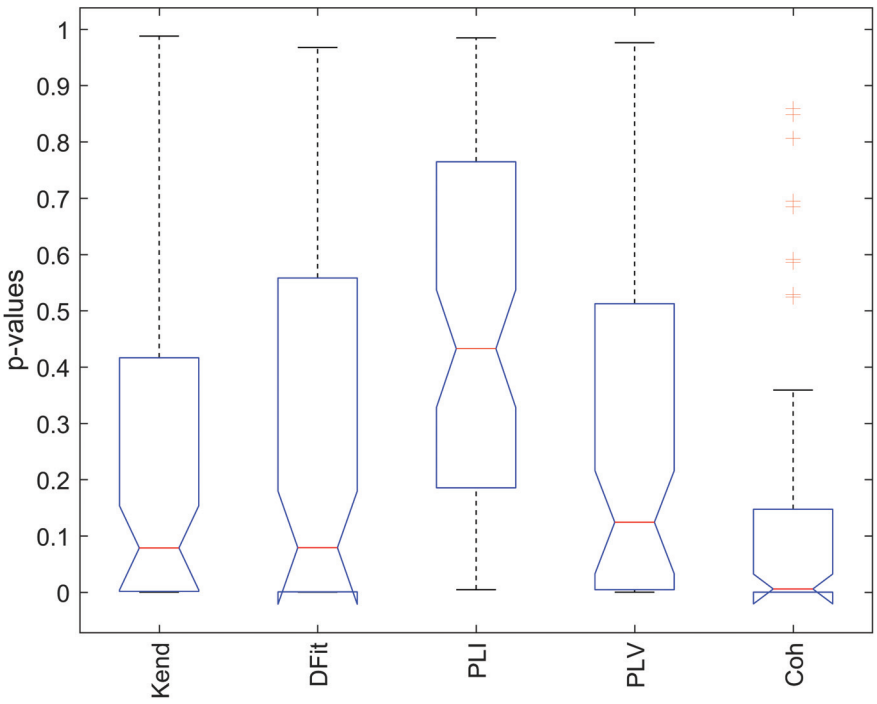


(b)

Figure 4.10. For time interval [0.2-4.1] sec a) average of logarithm of p-values for each task b) p-values boxplot of *both fist - both feet* task



(a)



(b)

Figure 4.11. For time interval [0.2-4.1] sec a) average of logarithm of p-values for each task at selected pairs b) p-values boxplot of *both fist - both feet* task at selected pairs

CHAPTER 5

DISCUSSION

In this study, 15 methods, 2 of which were proposed in this study, were evaluated in terms of their potential in capturing coherence between different EEG channel pairs. First of all, the computation time of these methods were analyzed on synthetic data. Correlation based methods, Spearman, Kendall and SCC coefficients and phase synchronization methods, PLV and PLI took less time than the others as expected, due to their simple structure. In addition, one of the proposed methods, data fitting based MI estimator compares well with the methods mentioned above, although it estimates mutual information. However, kNN based methods like Kraskov, REGO and the copula based method of Schweiser-Wolff showed a quadratically increasing time characteristic as the number of samples increased. Kraskov method took longer than UV parametrization for longer data in spite of its better performance on low sample sizes. Similarly, although the generalized synchronization index method performed well in low sample sizes, it consumed more and more time in search of neighbors even in modest embedding dimension values (in this study m was chosen as 10 and k as 1) as the sample size increased.

Secondly, a further time-accuracy analysis was implemented for information theoretic methods. As a result of the experiment on data with different degrees of dependence, sizes and probability density characteristics, data fitting, UV parametrization and Kraskov methods had better approximation to real mutual information values. As mentioned in Chapter 4, KGV had problems with estimating the MI value for highly dependent data albeit its independence detection ability was good. When the methods were analyzed along with computation times, data fitting based estimator had the best performance whereas Kraskov needed longer times for estimation, and KGV had worse error results.

A hierarchical clustering based evaluation experiment procedure was used in order to analyze the performance of methods as dependence or coherence measures for brain signals. In this test p-values were obtained for varying configurations to compare the significancy of the measures. Generally, better p values were observed for *both fist - both feet* task. Similarly, (Pfurtscheller and Neuper, 2001) declared that *right hand - both feet* task provides more accurate classification rather than *left hand - right hand*. In addition, statistical significance of the results improved with increasing signal duration.

Information theoretic methods had similar results for 1 sec and 2 sec epoch du-

rations except KGV: the p-value of KGV were relatively high. That may be cause of its gap at estimating the MI of highly dependent variables. Correlation based methods, Spearman, KCCA, Schweiser-Wolff and Kendall method showed good results for *both fist - both feet* task. Especially Kendall had the best p-value geometrical average for this task at 1 and 2 sec epoch durations. However, the standard deviation of these methods were worse than the mutual information methods of Kraskov, data fitting based estimator and REGO estimators. Finally, coherence and PLV methods had modest p-values. While they had larger standard deviations, especially coherence method's performance increased dramatically at 2 sec epoch duration.

PLI and iCOH methods achieved better discrimination between the tasks in the evaluation of 0.5 sec epoch although it was seen relatively high p values for other time interval experiments. In addition, for 1 sec and 2 sec time intervals, these methods showed better performances except for *both fist-both feet* task. On the other hand, although they had lower p-values than coherence and PLV methods, PLI and iCOH (especially PLI) had the lowest standard deviation among all measures in *both fist-both feet* task. There may be two reasons for that situation : they had the lowest standard deviation because they claim that they eliminate volume conduction effects and they had lower p-values because they miss zero-lag connectivities aiming to eliminate volume conduction effect.

Second experiment on EEG signals were applied on selected pairs. Similarly, the experiment that was implemented on 9 electrodes and 18 pairs, showed the best p-values for *both fist-both feet* task and 2 sec time interval where coherence method had the best p-value. In this experiment, information theoretic methods were relatively better than the other methods except for KGV again. Finally, Kendall method had close results to information theoretic methods. On the other hand, for this experiment PLI and iCOH methods had worse results for all tasks and time intervals. The main reason of the situation may be the small seach area of experiment. (Stam et al., 2007) stated that the local connections were absent for PLI and iCOH methods and the long distance interactions were dominant.

A further analysis using [0.2-4.1] time interval were applied on selected methods. Above mentioned experiment procedures were followed for this experiment again for 5 methods : coherence, data fitting based MI estimator, PLI, PLV and Kendall coefficient. For HiCl based procedure Kendall coefficient had best p-value and for the analysis on selected pairs coherence method had the best p-value. At the same time, coherence had the lowest standard deviation among all methods. On the other hand PLV, Kendall and data fitting based estimator methods gave similar p-values. Logarithmic p-value of PLI was lower in HiCl experiment in spite of its small standard deviation and worst p-value for

selected pairs again. Finally coherence method had a promising increase for *both fist-both feet imagery* task for experiment on selected pairs.

CHAPTER 6

CONCLUSION

In this thesis, fifteen methods have been used for functional connectivity or have the potential to be used to measure dependence for the evaluation of EEG motor/motor imagery signals. Two of these methods (data fitting based MI calculator and UV parametrization) were proposed as a novel technique for mutual information estimation.

In experiments of computation time and accuracy, a synthetic dataset was used. On the other hand for the statistical significance analysis Physionet EEG motor/motor imagery data was used. In computation time test, it was observed that information theoretic methods performed slower than the other methods. But the two new methods proposed in this thesis, seem to figure out this problem with their intuitive approaches. In addition, the accuracy test showed that these novel methods can be used as an estimator for mutual information.

Another novel technique proposed in this thesis is a hierarchical based procedure for the evaluation of statistical significancy. This procedure was used along with a similar statistical signifance test applied on the selected electrode pairs. In these tests, better p-values were obtained for *both fist - both feet* task. In addition, majority of the methods showed better performances with increasing time observed from the data. Exceptionally, PLI and iCOH methods proposed to avoid volume conduction showed better performances than the other methods on shorter times and imaginary tasks, and they had the lowest standard deviation although their geometric p-values were relatively high. In conclusion of tests applied with varying configuration, five of the methods were identified with high statistical significance and different dependence characteristics considering their computation times and accuracy performances: coherence, data fitting, Kendall, PLI and PLV. These methods should be considered as features of data and may be preferred to be used together rather than choosing one since these methods have different performance characteristics for observed data with different durations and different tasks.

For the future work, considering the gap of correlation based and information theoretic measures with time lagged dependencies, a time delay identification system may be utilized before applying these measures. Secondly, elucidated by the different characteristics of the measures' performances, a BCI classification system may be carried out by using multiple measures for feature extraction. And finally, standard deviations of mea-

sures' performances may lead us to a further subject specific analysis in order to detect the subject features that affect the performances.

REFERENCES

Adey, W. R., D. O. Walter, and C. Hendrix (1961). Computer techniques in correlation and spectral analyses of cerebral slow waves during discriminative behavior. *Experimental neurology 3*(6), 501–524.

Ahsan, M. R., M. I. Ibrahimy, O. O. Khalifa, et al. (2009). Emg signal classification for human computer interaction: a review. *European Journal of Scientific Research 33*(3), 480–501.

Akaho, S. (2006). A kernel method for canonical correlation analysis. *arXiv preprint cs/0609071*.

Allison, B., B. Graimann, and A. Gräser (2007). Why use a bci if you are healthy. In *ACE Workshop-Brain-Computer Interfaces and Games*, pp. 7–11.

Andrzejak, R. G., A. Kraskov, H. Stögbauer, F. Mormann, and T. Kreuz (2003). Bivariate surrogate techniques: necessity, strengths, and caveats. *Physical review E 68*(6), 066202.

Arnhold, J., P. Grassberger, K. Lehnertz, and C. E. Elger (1999). A robust method for detecting interdependences: application to intracranially recorded eeg. *Physica D: Nonlinear Phenomena 134*(4), 419–430.

Astolfi, L., F. Cincotti, D. Mattia, M. G. Marciani, L. A. Baccala, F. de Vico Fallani, S. Salinari, M. Ursino, M. Zavaglia, L. Ding, et al. (2007). Comparison of different cortical connectivity estimators for high-resolution eeg recordings. *Human brain mapping 28*(2), 143–157.

Bach, F. R. and M. I. Jordan (2002). Kernel independent component analysis. *Journal of machine learning research 3*(Jul), 1–48.

Bashashati, A., M. Fatourechi, R. K. Ward, and G. E. Birch (2007). A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. *Journal of Neural engineering 4*(2), R32.

Baule, G. M. (1963). Detection of the magnetic field of the heart. *American Heart Journal 66*, 95–96.

Belmonte, M. K., G. Allen, A. Beckel-Mitchener, L. M. Boulanger, R. A. Carper, and S. J. Webb (2004). Autism and abnormal development of brain connectivity. *Journal of Neuroscience 24*(42), 9228–9231.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289–300.

Berger, H. (1929). Über das elektrenkephalogramm des menschen. *Archiv für psychiatrie und nervenkrankheiten 87*(1), 527–570.

Bishop, C. M. (2006). Pattern recognition and machine learning (information science and statistics) springer-verlag new york. *Inc. Secaucus, NJ, USA*.

Bonita, J., L. Ambolode, B. Rosenberg, C. Cellucci, T. Watanabe, P. Rapp, and A. Albano (2014). Time domain measures of inter-channel eeg correlations: a comparison of linear, nonparametric and nonlinear measures. *Cognitive neurodynamics 8*(1), 1–15.

Borga, M., H. Knutsson, and T. Landelius (1997). Learning canonical correlations.

Bornot, J. M. S., K. Wong-Lin, A. L. Ahmad, and G. Prasad (2018). Robust eeg/meg based functional connectivity with the envelope of the imaginary coherence: Sensor space analysis. *Brain topography*, 1–22.

Cabeza, R., A. Kingstone, J. Raessens, et al. (2006). *Handbook of functional neuroimaging of cognition*. Mit Press.

Çağdaş, S. and B. Karaçali (2018). Novel techniques for model-free and fast computation of mutual information. In *2018 26th Signal Processing and Communications Applications Conference (SIU)*. IEEE.

Cheung, V., C. Cheung, G. McAlonan, Y. Deng, J. Wong, L. Yip, K. Tai, P. Khong, P. Sham, and S. Chua (2008). A diffusion tensor imaging study of structural

dysconnectivity in never-medicated, first-episode schizophrenia. *Psychological medicine 38*(6), 877–885.

Chicharro, D. and R. G. Andrzejak (2009). Reliable detection of directional couplings using rank statistics. *Physical Review E 80*(2), 026217.

Cohen, D. (1968). Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents. *Science 161*(3843), 784–786.

Daly, I., S. J. Nasuto, and K. Warwick (2012). Brain computer interface control via functional connectivity dynamics. *Pattern recognition 45*(6), 2123–2136.

Delorme, A. and S. Makeig (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods 134*(1), 9–21.

Delorme, A., S. Makeig, M. Fabre-Thorpe, and T. Sejnowski (2002). From single-trial eeg to brain area dynamics. *Neurocomputing 44*, 1057–1064.

Demuru, M., F. Fara, and M. Fraschini (2013). Brain network analysis of eeg functional connectivity during imagery hand movements. *Journal of integrative neuroscience 12*(04), 441–447.

Fallani, F. D. V., J. Richiardi, M. Chavez, and S. Achard (2014). Graph analysis of functional brain networks: practical issues in translational neuroscience. *Phil. Trans. R. Soc. B 369*(1653), 20130521.

Farwell, L. A. and E. Donchin (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology 70*(6), 510–523.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Human brain mapping 2*(1-2), 56–78.

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain connectivity 1*(1), 13–36.

Fukumizu, K., F. R. Bach, and A. Gretton (2007). Statistical consistency of kernel canonical correlation analysis. *Journal of Machine Learning Research 8*(Feb), 361–383.

Goldberger, A. L., L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation 101*(23), e215–e220.

Graimann, B., B. Allison, and G. Pfurtscheller (2010). Brain-computer interfaces revolutionizing human-computer interaction (the frontiers collection).

Hamedi, M., S.-H. Salleh, and A. M. Noor (2016). Electroencephalographic motor imagery brain connectivity analysis for bci: a review. *Neural computation 28*(6), 999–1041.

Hastie, T. J. and R. J. Tibshirani (1990). Generalized additive models, volume 43 of monographs on statistics and applied probability.

Haufe, S. and A. Ewald (2016). A simulation framework for benchmarking eeg-based brain connectivity estimation methodologies. *Brain topography*, 1–18.

Hero, A. O. and O. Michel (1998). Robust entropy estimation strategies based on edge weighted random graphs. In *Bayesian inference for inverse problems*, Volume 3459, pp. 250–262. International Society for Optics and Photonics.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on Neural Networks 10*(3), 626–634.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika 30*(1/2), 81–93.

Kirshner, S. and B. Póczos (2008). Ica and isa using schweizer-wolff measure of dependence. In *Proceedings of the 25th international conference on Machine learning*, pp. 464–471. ACM.

Kozachenko, L. and N. N. Leonenko (1987). Sample estimate of the entropy of a random vector. *Problemy Peredachi Informatsii 23*(2), 9–16.

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

Krauss, T. P., L. Shure, and J. N. Little (1994). Signal processing toolbox for use with matlab. Technical report, The Math Works.

Lachaux, J.-P., M. Chavez, and A. Lutz (2003). A simple measure of correlation across time, frequency and space between continuous brain signals. *Journal of neuroscience methods 123*(2), 175–188.

Lachaux, J.-P., A. Lutz, D. Rudrauf, D. Cosmelli, M. Le Van Quyen, J. Martinerie, and F. Varela (2002). Estimating the time-course of coherence between single-trial brain signals: an introduction to wavelet coherence. *Neurophysiologie Clinique/Clinical Neurophysiology 32*(3), 157–174.

Lachaux, J.-P., E. Rodriguez, J. Martinerie, and F. J. Varela (1999). Measuring phase synchrony in brain signals. *Human brain mapping 8*(4), 194–208.

Laureys, S., P. Peigneux, and S. Goldman (2002). Brain imaging. *Biological psychiatry*, 155–166.

Learned-Miller, E. G. and W. F. John III (2003). Ica using spacings estimates of entropy. *Journal of machine learning research 4*(Dec), 1271–1295.

Lee, Y.-Y. and S. Hsieh (2014). Classifying different emotional states by means of eeg-based functional connectivity patterns. *PloS one 9*(4), e95415.

Loboda, A., A. Margineanu, G. Rotariu, and A. M. Lazar (2014). Discrimination of eeg-based motor imagery tasks by means of a simple phase information method. *International Journal of Advanced Research in Artificial Intelligence 3*(10).

Maby, E., R. L. B. Jeannes, C. Liegeok-Chauvel, B. Gourevitch, and G. Faucon (2004). Analysis of auditory evoked potential parameters in the presence of radiofrequency fields using a support vector machines method. *Medical and Biological Engineering and Computing 42*(4), 562–568.

Mammone, N., C. Ieracitano, H. Adeli, A. Bramanti, and F. C. Morabito (2018).

Permutation jaccard distance-based hierarchical clustering to estimate eeg network density modifications in mci subjects. *IEEE Transactions on Neural Networks and Learning Systems*.

McKeeman, B. (2016). Matlab performance measurement. *MathWorks, Inc*.

Miller, E. G. (2003). A new class of entropy estimators for multi-dimensional densities. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, Volume 3, pp. III–297. IEEE.

Mitra, P. (2007). *Observed brain dynamics*. Oxford University Press.

Mitra, S. K. and Y. Kuo (2006). *Digital signal processing: a computer-based approach*, Volume 2. McGraw-Hill New York.

Mormann, F., K. Lehnertz, P. David, and C. E. Elger (2000). Mean phase coherence as a measure for phase synchronization and its application to the eeg of epilepsy patients. *Physica D: Nonlinear Phenomena 144*(3-4), 358–369.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Nolte, G., O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett (2004). Identifying true brain interaction from eeg data using the imaginary part of coherency. *Clinical neurophysiology 115*(10), 2292–2307.

Nunes, P. and R. Srinivasan (2006). Electric fields of the brain.

Odegard, T. N., E. A. Farris, J. Ring, R. McColl, and J. Black (2009). Brain connectivity in non-reading impaired children and children diagnosed with developmental dyslexia. *Neuropsychologia 47*(8-9), 1972–1977.

Oppenheim, A. V. and R. W. Schafer (2014). *Discrete-time signal processing*. Pearson Education.

Pál, D., B. Póczos, and C. Szepesvári (2010). Estimation of rényi entropy and mutual information based on generalized nearest-neighbor graphs. In *Advances in Neural*

Information Processing Systems, pp. 1849–1857.

Pascual-Marqui, R. D., D. Lehmann, M. Koukkou, K. Kochi, P. Anderer, B. Saletu, H. Tanaka, K. Hirata, E. R. John, L. Prichep, et al. (2011). Assessing interactions in the brain with exact low-resolution electromagnetic tomography. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 369*(1952), 3768–3784.

Pearson, K. (1920). Notes on the history of correlation. *Biometrika 13*(1), 25–45.

Peraza, L. R., A. U. Asghar, G. Green, and D. M. Halliday (2012). Volume conduction effects in brain network inference from electroencephalographic recordings using phase lag index. *Journal of neuroscience methods 207*(2), 189–199.

Pfurtscheller, G. and C. Neuper (2001). Motor imagery and direct brain-computer communication. *Proceedings of the IEEE 89*(7), 1123–1134.

Pfurtscheller, G. and C. Neuper (2006). Future prospects of erd/ers in the context of brain–computer interface (bci) developments. *Progress in brain research 159*, 433–437.

Póczos, B., S. Kirshner, and C. Szepesvári (2010). Rego: Rank-based estimation of rényi information using euclidean graph optimization. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 605–612.

Quiroga, R. Q., A. Kraskov, T. Kreuz, and P. Grassberger (2002). Performance of different synchronization measures in real data: a case study on electroencephalographic signals. *Physical Review E 65*(4), 041903.

Ranky, G. and S. Adamovich (2010). Analysis of a commercial eeg device for the control of a robot arm. In *Bioengineering Conference, Proceedings of the 2010 IEEE 36th Annual Northeast*, pp. 1–2. IEEE.

Redmond, C. and J. Yukich (1996). Asymptotics for euclidean functionals with power-weighted edges. *Stochastic processes and their applications 61*(2), 289–304.

Rényi, A. (1961). On measures of entropy and information. Technical report, HUNGARIAN ACADEMY OF SCIENCES Budapest Hungary.

Rosenblum, M. G., A. S. Pikovsky, and J. Kurths (1996). Phase synchronization of chaotic oscillators. *Physical review letters 76*(11), 1804.

Rulkov, N. F., M. M. Sushchik, L. S. Tsimring, and H. D. Abarbanel (1995). Generalized synchronization of chaos in directionally coupled chaotic systems. *Physical Review E 51*(2), 980.

Sakkalis, V. (2011). Review of advanced techniques for the estimation of brain connectivity measured with eeg/meg. *Computers in biology and medicine 41*(12), 1110–1117.

Sanei, S. and J. A. Chambers (2007). Eeg signal processing.

Schäfer, C., M. G. Rosenblum, H.-H. Abel, and J. Kurths (1999). Synchronization in the human cardiorespiratory system. *Physical Review E 60*(1), 857.

Schalk, G., D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw (2004). Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering 51*(6), 1034–1043.

Schweizer, B., E. F. Wolff, et al. (1981). On nonparametric measures of dependence for random variables. *The annals of statistics 9*(4), 879–885.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal 27*(3), 379–423.

Silfverhuth, M. J., H. Hintsala, J. Kortelainen, and T. Seppänen (2012). Experimental comparison of connectivity measures with simulated eeg signals. *Medical & biological engineering & computing 50*(7), 683–688.

Spearman, C. (2014). " general intelligence," objectively determined and measured. *American Journal of Psychology*.

Sporns, O. (2010). *Networks of the Brain*. MIT press.

Srinivasan, R., W. R. Winter, J. Ding, and P. L. Nunez (2007). Eeg and meg coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics. *Journal of neuroscience methods 166*(1), 41–52.

Stam, C. J., G. Nolte, and A. Daffertshofer (2007). Phase lag index: assessment of functional connectivity from multi channel eeg and meg with diminished bias from common sources. *Human brain mapping 28*(11), 1178–1193.

Stögbauer, H., A. Kraskov, S. A. Astakhov, and P. Grassberger (2004). Least-dependent-component analysis based on mutual information. *Physical Review E 70*(6), 066123.

Swarnkar, V., U. R. Abeyratne, and C. Hukins (2007). Inter-hemispheric asynchrony of the brain during events of apnoea and eeg arousals. *Physiological measurement 28*(8), 869.

Ter-Pogossian, M. M., M. E. Raichle, and B. E. Sobel (1980). Positron-emission tomography. *Scientific American 243*(4), 170–181.

Thatcher, R. W., C. J. Biver, D. M. North, and C. R. W. Thatcher (2004). Eeg and brain connectivity: A tutorial. *Unpublished manuscript*.

Tummala, N., K. Venkatasubramanian, and V. Umamaheswari (2018). Embedded system for classification of upper limb movement during action using eeg. In *Intelligent Embedded Systems*, pp. 241–250. Springer.

Vajda, I. (1989). *Theory of statistical inference and information*, Volume 11. Kluwer Academic Pub.

van Diessen, E., S. J. Diederen, K. P. Braun, F. E. Jansen, and C. J. Stam (2013). Functional and structural brain networks in epilepsy: what have we learned? *Epilepsia 54*(11), 1855–1865.

Varela, F., J.-P. Lachaux, E. Rodriguez, and J. Martinerie (2001). The brainweb: phase synchronization and large-scale integration. *Nature reviews neuroscience 2*(4), 229.

Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54–59.

Westfall, P. H. and S. S. Young (1993). Resampling-based multiple testing: Examples and methods for p-value adjustment (wiley series in probability and statistics).

Wieczorkowski, R. and P. Grzegorzewski (1999). Entropy estimators-improvements and comparisons. *Communications in Statistics-Simulation and Computation 28*(2), 541–567.

Zhu, D., J. Bieger, G. G. Molina, and R. M. Aarts (2010). A survey of stimulation methods used in ssvep-based bcis. *Computational intelligence and neuroscience 2010*, 1.

APPENDIX A

BACKGROUND

A.1. Independent Component Analysis (ICA)

Independent component analysis is a feature extraction method that aims to reconstruct the independent source signals from the linear mixtures of these source signals. The model of linear mixtures is:

$$x(t) = As(t) \tag{A.1}$$

where $s(t)$ represents the source signals vector, x is the mixture vector and A is the mixing matrix. ICA aims to find the demixing matrix W (inverse of A).

$$s(t) = Wx(t) \tag{A.2}$$

In search of matrix W without knowing A, W is decomposed into prewhitening matrix V and rotation matrix R.

$$W = RV \tag{A.3}$$

After finding prewhitening matrix, the only challenge is to find the proper rotation angle that gives independent sources. At this step, the performance of each angle is evaluated by the contrast functions. These contrast functions may measure either independency or nongaussianity of reconstructed signals as the objective of ICA method. Some of the contrast functions that measures independency such as KCCA and REGO are also analyzed in this study as the dependence measure for brain connectivity.

APPENDIX B

MATERIAL AND METHODS

B.1. Mathematical Background of Zero-phase Filtering

Let $H(e^{jw})$ be the frequency response of any IIR filter and $X(e^{jw}$ be the representation of a signal. The output signal after filtering the signal by this system is defined as

$$V(e^{jw}) = H(e^{jw})X(e^{jw}) \tag{B.1}$$

By following the zero-phase filter, $U(e^{jw})$ is carried out after the time-reverse operation:

$$U(e^{jw}) = V^*(e^{jw}) \tag{B.2}$$

In the next $W(e^{jw})$ is calculated by filtering this time-reversed signal one more time. The relation of this signal with the original signal can be carried out by

$$W(e^{jw}) = H(e^{jw})U(e^{jw}) = H(e^{jw})H^*(e^{jw})X^*(e^{jw}) = |H(e^{jw})|^2X^*(e^{jw}) \tag{B.3}$$

In order to mathematical expression, it can be easily seen that, filter causes a change only on the magnitude of the signal. However, the output is still in the time-reversed way. Because of that, in the last step the signal is reversed back in time and zero-phase filter is completed:

$$Y(e^{jw}) = W*(e^{jw}) = |H(e^{jw})|^2X(e^{jw}) \tag{B.4}$$

B.2. Hilbert Transform

Hilbert Transform of any signal $g(t)$ is calculated by a convolution operation defined as

$$\begin{aligned}\hat{x} &= \frac{1}{\pi t}\int \frac{g(t')}{t-t'}dt' \\ &= \frac{1}{\pi t}\int \frac{g(t')}{t-t'}dt'\end{aligned} \tag{B.5}$$

Fourier transform of the convolution operator is described as

$$\frac{1}{\pi t} \rightarrow jsgn(f) \tag{B.6}$$

which means that Hilbert transform causes a 90 degree phase shift on $g(t)$. Because of that the it is also called *phase shifter*.

B.3. Wavelet Transform

Short time Fourier transform (STFT) helps detecting the change of frequency components of any signal $x(t)$ in time by applying Fourier transform in time bins as

$$X(w,t) = \int_{-\infty}^{\infty} x(t)[h(t-\tau)e^{-iw\tau}]d\tau \tag{B.7}$$

where $h(t)$ denotes the time bin (window) that the transform is applied. For STFT method, the window length parameter should be chosen carefully the window length affects the time and frequency resolution, and for high frequency components, frequency resolution is important whereas the time resolution matters more for low frequency components. Because of that, applying an adaptive window length for corresponding frequency component, wavelet transform provides a better transformation as

$$W(\tau,s) = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{|s|}}\psi * \left(\frac{t-\tau}{s}\right)d\tau \tag{B.8}$$

where τ denotes the position of the wavelet in time domain and the s is the scale parameter. $*$ denotes the complex conjugate of wavelet function ψ.

B.4. Variance Estimation for Data Fitting MI Calculator

Running mean smoothing (RMS) and polynomial regression methods are used for conditional variance $(\sigma^2(Y|X))$ estimation. A running mean smoother is a moving average filter actually. It proves to see the trend of the variable (conditional variance in

this case). The smoother output for a sample is the average of of the sample itself and its neighborhoods. But how big to take the neighborhood affects the smoother. If large neighborhood is chosen this causes low variance and high bias in estimator. On the other hand, small neighborhood may cause high variance and overfitting problem.

The other method for estimation, polynomial regression aims to fit the data to a polynomial equation. For linear regression the equation is:

$$\hat{\sigma}^2(Y|X) = \alpha + \beta x \tag{B.9}$$

Regression is implemented by finding the coefficients α, β that give the minimum error on fitting. For polynomial regression, number of predictor or degree of polynomial affects the output similar to RMS. Choosing low degree causes bias whereas high degree causes overfitting problem.

In this study, parameters for both methods are chosen by a cross validation based algorithm:

1. Define the candidate values.

   Neighborhood of RMS (ratio to total number of samples): $\{$0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4 0.5 0.6, 0.7$\}$

   Degree of polynomial regression : $\{$1, 2, 3, 4, 5$\}$

2. Sort the data by X. Divide data into two groups as even and odd ranked samples.

3. For each candidate value, use odd ranked samples for estimation and even ranked samples to calculate mean square error (mse).

4. Choose the value with smallest mse as parameter.

B.5. Relation of Copula and Mutual Information

As defined in Schweizer-Wolff and REGO methods, cumulative distribution of two univariate variables are defined as new variables $U = F_x(X)$ and $V = F_y(X)$ where the copula function is $C(u, v) = F(F_x^{-1}(u), F_y^{-1}(v))$. The copula density function $c(u, v)$ can be extracted easily:

$$c(u,v) = \frac{\partial^2 C(u,v))}{\partial u \partial v} = \frac{\partial^2 F(x,y))}{\partial F_x(x) \partial F_y(y)} = \frac{\frac{\partial^2 F(x,y))}{\partial x \partial y}}{\frac{\partial F_x(x)}{\partial x}\frac{\partial F_y(y)}{\partial y}} \tag{B.10}$$

$$c(u,v) = \frac{f(x,y)}{f_x(x)f_y(y)} \tag{B.11}$$

To reformulate the mutual information by using copula density function, standard formula is multiplied by $\frac{\partial F_x(x)\partial F_y(y)}{\partial x \partial y}$ and divide by $f_x(x)f_y(y)$ by using the relation $f_x(x) = \frac{\partial F_x(x)}{\partial x}$:

$$\begin{aligned} I(X,Y) &= \int \frac{f(x,y)}{f_x(x)f_y(y)} \log \frac{f(x,y)}{f_x(x)f_y(y)} \frac{\partial F_x(x)\partial F_y(y)}{\partial x \partial y} dxdy \\ &= \int \frac{f(x,y)}{f_x(x)f_y(y)} \log \frac{f(x,y)}{f_x(x)f_y(y)} dF_x(x)dF_y(y) \end{aligned} \tag{B.12}$$

By using the equation B.11 and place the U and V variables, new mutual information equation based on copula density is easily written and the relation is proven:

$$I(X,Y) = \int c(u,v) \log c(u,v) dudv \tag{B.13}$$

APPENDIX C

EXPERIMENTAL METHOD

C.1. Timing Test Procedure on Windows Powershell

To manage processes in Windows PC, a series of code in powershell interface were used in administrator mode. First of all running processes were reached by

```
$instances = Get-Process
```

code. After that two lines of code were used to allocate core 8 for Matlab, set the priority of Matlab as high and the others below normal:

```
foreach ($i in $instances)
{
if($i.ProcessName -eq "matlab") {$i.ProcessorAffinity=128 }
else
{$i.ProcessorAffinity=127}
}

foreach ($i in $instances)
{
if($i.ProcessName -eq "matlab")
{$i.PriorityClass = 'high'}
else
{$i.PriorityClass = 'BelowNormal'}
}
```

At last, all processes other than Matlab and powershell interface were killed:

```
foreach ($i in $instances)
{
if($i.ProcessName -ne "conhost"
   -and $i.ProcessName -ne "powershell"
   -and $i.ProcessName -ne "matlab"  )
{kill -id $i.id}}
```